ROYAUME DU MAROC
Université Mohammed V
- RABAT -
جامعة محمد الخامس
- الرباط -
Faculté des sciences
كلية العلوم
CENTRE D'ÉTUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

N° d'ordre : 3901

THÈSE

En vue de l'obtention du : ***DOCTORAT***

Structure de Recherche : Intelligent Processing Systems and Security (IPSS)
Discipline : Informatique
Spécialité : Intelligence Artificielle

Présentée et soutenue le : 16/02/2024
Par :

Hajar Zankadi

AI-driven approach for learner 's profile personalization within MOOC integrating social media information

JURY

Abderrafiaa KOUKAM	PES, Université de Bourgogne Franche Comté, France	Président
Fouzia OMARY	PES, Faculté des Sciences, Université Mohammed V de Rabat	Rapportrice / Examinatrice
Faouzia BENABBOU	PES, Faculté des Sciences, Hassan II University, Casablanca	Rapportrice / Examinatrice
Nour El Houda CHAOUI	PES, Université Ibn Tofail, Kenitra	Rapportrice/ Examinatrice
Hassan ECHOUKAIRI	PH, Faculté des Sciences, Université Mohammed V de Rabat	Examinateur
Abderrahmane EZZAHOUT	PH, Faculté des Sciences, Université Mohammed V de Rabat	Examinateur
Imane Hilal	PH, Ecole des Sciences de l'Information (ESI) de Rabat	Co-directrice de thèse
Abdellah IDRISSI	PES, Faculté des Sciences, Université Mohammed V de Rabat	Directeur de thèse

Année Universitaire: 2022/2023

DEDICATION

I would like to take this opportunity to let my family know that I am grateful for their constant support:

- *To you **Hajar**, thank you for not giving up, for working hard, for being patient and considerate to achieve your goal and defend your thesis.*

- *To my husband, **Othmane**, to who I owe deepest gratitude. I am forever thankful for the unconditional love and support throughout this challenging journey. You were always there, wherever and whenever that was. Thank you for your support, patience and encouragement during my ups and downs.*

- *To my son, **Kossai** who keeps me grounded, remind me of what is important in life, who is my source of motivation to not giving up.*

- *To my mom and dad, **Saidia** and **Ismail**, for their endless love, throughout my life. I appreciate you both for giving me the will to aim high and follow my dreams. You have always been supportive of me, prayed for me, and had faith in me.*

- *To my sisters, **Kawtar** and **Meryam**, whose constant love and support provided me with strength.*
- *To my aunts, **Fatima**, **Zahra** and **Rekia**, to whom I express my deepest gratitude and my heartfelt thanks.*

- *To my step mom and step dad, **Saidia** and **Abderahman**, who were there by my side, offering me love and support.*

- *To the rest of my family and to my friends, thank you for your understanding and encouragement in my difficult moments.*

ACKNOWLEDGEMENT

The work presented in this thesis has been done in the Computer Science Laboratory of Mohammed V University, Faculty of Sciences of Rabat, more specifically, in the research structure: Intelligent Processing Systems and Security (IPSS).

I would like to start by expressing my gratitude to Mr. **Abdellah IDRISSI**, Director of my thesis for having accepted to direct this thesis. It has been a genuine honor working with them.

I would like to thank as well, Mrs. **Imane HILAL** for having accepted to direct this thesis. It has been a genuine honor working with them. Thank you for your commitment, guidance, and support throughout the course of my research. I appreciate the great opportunity you provided for me to experience and the chances you gave me to advance professionally.

I thank also Mr. **Abderrafiaa KOUKAM,** PES at the University of Bourgogne Franche-Comté, France for his time and interest to evaluate our work.

I am also grateful to Mrs. **Fouzia OMARY**, head of the research structure IPSS and PES at the Faculty of Sciences, University Mohammed V - Rabat, for her input and for her interest in our work. Additionally, I would also thank her for her support, help and encouragement during this Ph.D. journey.

I would also like to express my gratitude to Mrs. **Faouzia BENABBOU**, PES at the Faculty of Sciences, Hassan II University – Casablanca, for agreeing to participate in this jury as rapporteur and for reading my dissertation.

My sincere thanks also go to Mrs. **Nour El Houda CHAOUI**, PES at Université Ibn Tofail Kenitra, for agreeing to participate in this jury as rapporteur and for reviewing my dissertation.

I express my gratitude to Mr. **Hassan ECHOUKAIRI,** PH at the Faculty of Sciences, University Mohammed V - Rabat, for having accepted to evaluate our work and for their time, feedback and interest in our thesis.

I also extend my thanks to Mr. **Abderrahmane EZZAHOUT**, PH at the Faculty of Sciences, University Mohammed V - Rabat , agreeing to participate in this jury and for having accepted to evaluate our work.

I would like to extend my warmest thanks to all my professors from INPT, who impacted and inspired me.

Résumé

L'apprentissage en ligne a révolutionné l'éducation en offrant des possibilités d'apprentissage accessibles et flexibles. Les apprenants peuvent accéder à diverses ressources et bénéficier d'expériences personnalisées, grâce aux avancées technologiques de pointe. Il favorise l'apprentissage tout au long de la vie et le développement professionnel, tout en encourageant la collaboration et le travail en réseau entre les apprenants. L'apprentissage en ligne s'est avéré rentable et résilient en cas de perturbation COVID 19.

Par conséquent, l'apprentissage en ligne est devenu une partie intégrante de l'éducation moderne, offrant des expériences d'apprentissage pratiques, inclusives et engageantes pour les apprenants du monde entier. Cependant, les taux d'abandon dans les MOOC restent un défi en raison de facteurs tels que le manque de motivation et d'engagement continus, la diminution des interactions sociales et les recommandations de contenu impersonnalisées.

Pour résoudre ces problèmes, notre approche de recherche explore l'intersection des MOOC et des médias sociaux en utilisant des approches basées sur l'IA. Nous visons à enrichir l'expérience des apprenants au sein des MOOC en tirant parti des techniques d'IA, telles que les ontologies, la modélisation des sujets, le traitement du langage naturel (NLP) et les systèmes de recommandation. Par conséquent, l'intégration des médias sociaux dans les plateformes de MOOC stimulera l'engagement et la personnalisation. Pour mettre en œuvre notre approche, nous avons d'abord utilisé l'ontologie comme base d'IA pour le contenu de la base de connaissances, afin de développer une ontologie globale et complète qui représente les profils sociaux des apprenants. Cette ontologie intègre de nombreuses techniques d'ingénierie telles que la cartographie et la fusion d'ontologies pour englober un large éventail d'attributs, y compris des informations démographiques, des intérêts, des préférences, des styles d'apprentissage, des connexions sociales, etc. Notre objectif principal est de créer une vue holistique de chaque apprenant, permettant une expérience d'apprentissage plus adaptée et personnalisée.

Sur la base de l'ontologie précédente, nous proposons un modèle de sujet de cours (CTM) utilisant des techniques de modélisation de sujet et de NLP pour identifier et détecter les cours d'intérêt des apprenants à partir de leurs interactions spontanées dans les médias sociaux, en particulier Twitter. Nous mettons en œuvre de nombreuses méthodes d'apprentissage automatique non supervisées telles que l'allocation de Dirichlet latent (LDA), l'analyse sémantique latente (LSA) et BERTopic après avoir appliqué un pipeline NLP sur les tweets partagés par les apprenants. L'évaluation des modèles révèle que le modèle BERTopic est plus performant sur l'ensemble de données mis au rebut et leurs résultats sont utilisés pour générer le modèle de sujet de cours. Dans cette étape, nous obtenons des informations plus approfondies sur les préférences des apprenants afin de nous aligner plus étroitement sur leurs besoins et aspirations individuels.

Les résultats des deux contributions (l'ontologie des profils sociaux (SPOnt) et le modèle des thèmes de cours (CTM)) pourraient être utilisés pour enrichir le processus de recommandation utilisé aujourd'hui dans l'environnement MOOC.

Mots clés : ontologies, NLP, modélisation thématique, MOOC, médias sociaux, profil social.

ABSTRACT

E-learning has revolutionized education by providing accessible and flexible learning opportunities. Learners can access diverse resources and benefit from personalized experiences, thanks to cutting edges advances in technologies. It promotes lifelong learning and professional development, while fostering collaboration and networking among learners. E-learning has proved to be cost-effective and resilient during COVID 19 disruption. Therefore, e-learning has become an integral part of modern education, offering convenient, inclusive, and engaging learning experiences for learners worldwide.

E-learning, particularly Massive Open Online Courses (MOOCs), offers free access to educational content, expanding opportunities for self-directed learning and personal growth. However, dropout rates in MOOCs remain a challenge due to factors such as lack of continuous motivation and engagement, less social interactions, and impersonalized content recommendations.

To address these issues, our research approach explores the intersection of MOOCs and social media using AI-driven approach. We aim to enrich the learners' experience within MOOCs by leveraging AI techniques, such as ontology, natural language processing (NLP), topic modeling, and recommender systems. Therefore, the integration of social media into MOOC platforms will boost engagement and personalization.

To implement our approach, we first utilized ontology as an AI foundation for knowledge base content, to develop a global and comprehensive ontology that represents the learners' social profiles. This ontology incorporates many engineering techniques such as ontology mapping and merging to encompass a wide range of attributes, including demographic information, interests, preferences, learning styles, social connections and etc., gathered from both MOOC and social media platforms profiles. Our main purpose is to create a holistic view of each learner, enabling a more tailored and personalized learning experience.

Based on the previous ontology, we propose a Course Topic Model (CTM) using topic modeling and NLP techniques to identify and detect learners' course of interest from their spontaneous interaction in social media, in particular Twitter. We implement many unsupervised machine learning methods such as Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA) and BERTopic after applying NLP pipeline on the tweets shared by the learners. The evaluation of the models reveals that BERTopic model performed better on the scrapped dataset and their results are used to generate the course topic model. In this step, we gain deeper insights into the learner preferences to align more closely with their individual needs and aspirations.

The output of both contributions (Social profile ontology (SPOnt) and Course Topic Model (CTM)) could be used to enrich the recommendation process that is used nowadays in MOOC environment.

Keywords: Ontologies, NLP**,** Topic Modeling, MOOCs, social media, Social Profile.

الملخص

يعتبر تقدم التعلم الإلكتروني ثورة في التعليم من خلال توفير فرص تعلم مرنة ومتاحة. يمكن للمتعلمين الوصول إلى مصادر متنوعة والاستفادة من تجارب تعلم مخصصة، بفضل التطورات الحديثة في التكنولوجيا. يعزز التعلم الإلكتروني التعلم مدى الحياة والتطوير المهني، بينما يعزز التعاون والتواصل بين المتعلمين. أثبت التعلم الإلكتروني كفاءته ومرونته خلال COVID 19، لذلك، أصبح التعلم الإلكتروني جزءًا أساسيًا من التعليم الحديث، يقدم تجارب تعلم مريحة وشاملة للمتعلمين في جميع أنحاء العالم.

يقدم التعلم الإلكتروني، وبخاصة المساق التعليمي هائل الالتحاق(MOOCs) ، وصولًا مجانيًا إلى المحتوى التعليمي، مما يوسع الفرص للتعلم الذاتي والنمو الشخصي. ومع ذلك، تظل نسب الانقطاع في دورات MOOCs تحديًا نظرًا لعوامل مثل نقص التحفيز المستمر والانخراط، وقلة التفاعل الاجتماعي، وانعدام التوصيات المخصصة.

لمعالجة هذه المشكلات، يستكشف نهج بحثنا التقاطع بين المساق التعليمي هائل الالتحاق (MOOCs) ووسائل التواصل الاجتماعي باستخدام نهج يعتمد على الذكاء الاصطناعي. نهدف إلى إثراء تجربة المتعلمين داخل دورات المساق التعليمي هائل الالتحاق من خلال استخدام تقنيات الذكاء الاصطناعي، مثل الأنطولوجيا، ومعالجة اللغة الطبيعية(NLP) ، ونمذجة المواضيع، وأنظمة التوصية. لذلك، سيعزز دمج وسائل التواصل الاجتماعي في منصات دورات الانترنت المفتوحة الضخمة الانخراط والتخصيص.

لتنفيذ نهجنا، استخدمنا أولاً الأنطولوجيا كأساس للذكاء الاصطناعي لمحتوى قاعدة المعرفة، لتطوير أنطولوجيا شاملة عالمية تمثل ملفات تعريف المتعلمين الاجتماعية. تضم هذه الأنطولوجيا العديد من التقنيات الهندسية مثل بناء الأنطولوجيا ودمجها لتشمل مجموعة واسعة من السمات، بما في ذلك المعلومات الديموغرافية، والاهتمامات، والتفضيلات، وأساليب التعلم، والاتصالات الاجتماعية وما إلى ذلك، التي يتم جمعها من ملفات تعريف المتعلمين على المساق التعليمي هائل الالتحاق ووسائل التواصل الاجتماعي. هدفنا الرئيسي هو إنشاء رؤية شاملة لكل متعلم، مما يتيح تجربة تعلم أكثر تخصيصًا وتخصيصًا.

بناءً على الأنطولوجيا السابقة، نقترح نموذج الموضوع Topic Modeling باستخدام تقنيات نمذجة الموضوع ومعالجة اللغة الطبيعية لتحديد واكتشاف دورات الاهتمام للمتعلمين من تفاعلهم العفوي في وسائل التواصل الاجتماعي، بشكل خاص تويتر. نقوم بتنفيذ العديد من أساليب تعلم الآلة غير المشرفة مثل (LDA)، وتحليل المعنى اللاتني (LSA) و BERTopic بعد تطبيق معالجة اللغة الطبيعية على التغريدات المشاركة من قبل المتعلمين. تُظهر تقييم النماذج أن نموذج BERTopic يعمل بشكل أفضل على مجموعة البيانات المستخرجة ويتم استخدام نتائجها لإنشاء نموذج موضوع. في هذه الخطوة، نحصل على رؤى أعمق في تفضيلات المتعلمين لمواءمة أكثر مع احتياجاتهم وتطلعاتهم الفردية.

يمكن استخدام نتائج كل من المساهمات أنطولوجيا ملفات تعريف الشخصية الاجتماعية (SPOnt) ونموذج موضوع الدورة ((CTM)لإثراء عملية التوصية التي تستخدم حاليًا في بيئة دورات الانترنت المفتوحة الضخمة.

الكلمات الرئيسية: الأنطولوجيات، معالجة اللغة الطبيعية، نمذجة المواضيع المساق التعليمي هائل الالتحاق ، وسائل التواصل الاجتماعي، الملف الشخصي الاجتماعي.

RESUME DETAILLE

1. Introduction:

L'apprentissage en ligne est un sujet de discussion depuis les années 1990, mais son évolution a attiré une attention considérable, en particulier pendant la pandémie de COVID 19, où le passage à l'apprentissage en ligne est maintenant nécessaire (Martin et al. 2022). L'apprentissage en ligne est défini comme "l'acquisition de connaissances par le biais de technologies et de médias électroniques (Kim and Park 2021) (Milićević et al. 2021)" dans le but de fournir des ressources d'apprentissage plus largement accessibles, libérées des contraintes d'espace, de temps et de distance, aux apprenants du monde entier (Al-Fraihat, Joy, and Sinclair 2020) (Cidral et al. 2018). Cela a fait de l'apprentissage en ligne l'un des secteurs les plus appréciés de l'économie mondiale et l'une des approches éducatives les plus utilisées au 21e siècle (Sarwari et al. 2022).

En tant qu'outil d'apprentissage en ligne très efficace, adaptable et à la pointe de la technologie, les MOOC ont attiré l'attention. Leurs avantages remarquables en termes d'ouverture, d'accès illimité au matériel, de partage des programmes et de faible coût ont contribué à leur croissance rapide. Ils ont excellé dans les disciplines liées aux sciences humaines, à la gestion d'entreprise, à la santé et aux sciences, augmentant les inscriptions dans le monde entier et l'engagement des apprenants via des portails web dédiés, quel que soit leur lieu de résidence. Ils se sont révélés être un outil d'enseignement et d'apprentissage efficace et une alternative aux techniques conventionnelles d'apprentissage en face à face. Les faibles taux d'achèvement des cours et les taux élevés d'abandon ont toutefois un impact significatif sur la persévérance des participants aux MOOC.

De nombreux facteurs influencent les décisions des apprenants à tous les stades de l'acceptation des MOOC, notamment le sentiment d'isolement et le manque d'interaction, le manque de motivation permanente, les caractéristiques de l'apprenant, les caractéristiques des MOOC et l'expérience de l'apprenant (Badali et al. 2022) (Gupta and Maurya 2022).

D'autre part, les médias sociaux sont de plus en plus utilisés en raison de leur large diffusion (Facebook, Twitter, YouTube, Instagram, TikTok, etc.).). La connectivité, l'interactivité et le contenu généré par les utilisateurs sont les caractéristiques qui les définissent. En outre, les médias sociaux ont un impact significatif sur le partage d'informations par le biais d'une communication interactive qui reflète les pensées, les sentiments et les actions des utilisateurs. Par conséquent, l'utilisation des médias sociaux et leurs effets sur la socialisation constituent une source fiable pour recueillir et analyser les préférences et les intérêts des utilisateurs.

Dans ce contexte, l'apprentissage en ligne soutenu par les médias sociaux peut renforcer l'engagement et la coopération en réunissant des personnes ayant des objectifs et des intérêts similaires, en encourageant la participation active, en augmentant la variété et l'accessibilité de l'apprentissage, et en unifiant et partageant le contenu (Mak, Poon, and Chiu 2022). Par conséquent, l'inclusion d'informations sur les médias sociaux et l'intégration de leurs caractéristiques dans les processus d'enseignement et d'apprentissage des MOOC peuvent avoir un impact positif sur les apprenants. Selon (Salloum et al. 2021), les médias sociaux ont modernisé l'apprentissage en ligne en introduisant de nouvelles fonctionnalités telles que la réaction, le commentaire, la motivation ou la création de groupes. Ils ont un impact bénéfique majeur sur l'utilité perçue et la facilité d'utilisation perçue des environnements d'apprentissage en ligne (Troussas, Krouska, and Sgouropoulou 2021). Offrir aux apprenants un apprentissage en ligne social qui répond à leurs intérêts et à leurs préférences peut ainsi renforcer leur rôle dans la sociabilité, le sentiment d'appartenance à une communauté et la satisfaction à l'égard du cours, et donc encourager l'engagement (Yılmaz and Yılmaz 2022).

Les MOOC ont gagné en popularité depuis 2008 dans le domaine de l'enseignement en ligne. Les universités et les établissements d'enseignement supérieur proposent des cours en ligne flexibles et de qualité qui peuvent être suivis par des apprenants ayant des besoins et des préférences d'apprentissage variés. Sans restriction concernant la taille des classes, les compétences nécessaires ou les conditions d'inscription, les MOOC permettent aux apprenants de s'engager dans des communautés en ligne ouvertes et publiques. Ils sont souvent gratuits, sauf lorsqu'un certificat d'achèvement est délivré (Jin 2021; Goopio and Cheung 2021).

Pendant la pandémie de COVID 19, les MOOC ont fait l'objet d'une attention accrue et sont désormais considérés comme une méthode d'apprentissage très efficace et adaptable. Récemment, plus de 2,8 000 cours ont été ajoutés et 16,3 000 MOOC ont été annoncés ou lancés par environ 950 universités dans le monde, ce qui démontre la forte demande de MOOC (Er-Rafyg et al., n.d.). Selon les données rendues publiques par Class Central, les inscriptions aux MOOC ont augmenté d'un tiers, atteignant 180 millions d'apprenants dans le monde (à l'exclusion de la Chine). Les principaux fournisseurs de MOOC, Coursera, edX, Udacity et FutureLearn, ont lancé plus de 360 micro-crédits et 19 diplômes en ligne en 2020. Nombre d'entre eux ont proposé des certificats complémentaires.

Cependant, plusieurs problèmes rencontrés par les MOOC ont été mis en évidence, notamment les taux élevés d'abandon et les faibles taux de rétention qui se produisent dans les premières étapes de l'apprentissage (Chen et al. 2019). L'abandon dans les MOOC se produit lorsque les apprenants ne parviennent pas à terminer le cours (Dass, Gary, and Cunningham 2021). Selon (Badali et al. 2022), environ 90 % des apprenants inscrits abandonnent avant la fin du cours, et le taux de rétention pour les MOOC se situe entre 3 et 15 %.

Plusieurs facteurs, dont le manque de motivation et d'engagement continus, la diminution de l'interaction, les caractéristiques de l'apprenant, les caractéristiques du MOOC et l'expérience de l'apprenant, ont un impact sur les décisions prises par les apprenants à tous les stades de l'acceptation du MOOC (Badali et al. 2022) (Gupta and Maurya 2022) (Panagiotakopoulos et al. 2021) (Ji, Park, and Shin 2022).

Nous nous concentrons sur les caractéristiques et la situation sociale de l'apprenant, car la décision d'abandonner dépend davantage du niveau d'intégration sociale et intellectuelle, comme l'expliquent (Lee and Choi 2011). Selon l'auteur, la probabilité qu'un apprenant abandonne ses études augmente s'il a du mal à s'intégrer dans la communauté sociale ou à adhérer aux normes académiques de l'établissement d'enseignement supérieur. Dans le même ordre d'idées, d'autres auteurs ont souligné l'importance des caractéristiques sociales telles que la présence sociale et l'engagement pour influencer les intentions des apprenants de terminer leurs cours (Rosé et al. 2014); (Yang, Wen, and Rose 2014); (Zheng et al. 2015) ;(Wang et al. 2019); (Zhang et al. 2016). D'autres (Gütl et al. 2014); (Khalil and Ebner 2014); (Shapiro et al. 2017) ont attribué le taux d'abandon élevé dans les MOOC à certaines caractéristiques personnelles ou au manque de motivation (Xiong et al. 2015); (Khalil and Ebner 2014).

L'engagement et la satisfaction de l'apprenant sont des facteurs essentiels pour refléter l'efficacité des MOOC et sont tous deux considérés comme des moteurs cruciaux de la réussite de l'apprentissage (Han, Geng, and Wang 2021) (Alqurashi 2016). Selon (Crane and Comley 2021), l'apprentissage social a démontré qu'il permettait d'accroître la satisfaction des apprenants en réduisant le sentiment d'isolement et le manque d'interactions interpersonnelles. Dans (Liu et al. 2022), les auteurs révèlent que l'engagement et la satisfaction des apprenants sont améliorés lorsque les apprenants ont à la fois une motivation intrinsèque (intérêt) et une motivation extrinsèque (valeur perçue des connaissances).

L'apprentissage social en ligne fait référence au développement de plateformes d'apprentissage en ligne dotées de fonctions de médias sociaux ou de sites de médias sociaux utilisés à des fins éducatives (Troussas, Krouska, and Sgouropoulou 2021). L'utilisation des médias sociaux dans l'apprentissage en ligne permet à l'apprentissage de se dérouler sans être limité par des lieux physiques et de diverses manières créatives, y compris l'interaction sociale, les collaborations en ligne, l'interface multimédia personnalisée et conviviale qui offre un niveau élevé de participation, ainsi que la coopération et le contact entre les utilisateurs (Krouska, Troussas, and Virvou 2019) (Hajli et al. 2013). En outre, les médias sociaux peuvent grandement améliorer l'engagement des apprenants dans le processus d'apprentissage en favorisant l'interaction entre les apprenants et les enseignants, le retour d'information sur les supports de cours sous forme de likes et de commentaires, la motivation des apprenants sous forme de notifications, la formation de groupes d'apprenants cohésifs pour les tâches collaboratives, entre autres choses (Alalwan et al. 2019). Des études ont montré que les médias sociaux offrent des opportunités avantageuses pour l'enseignement et l'apprentissage à la lumière des approches constructivistes promues. Cela signifie que l'apprentissage se produit lorsque les apprenants contribuent activement au processus de construction des connaissances plutôt que de se contenter d'absorber des informations (Gray 1997). Ainsi, les médias sociaux ont été mis en évidence comme ayant la capacité de stimuler la mise en réseau et l'engagement entre les enseignants et les apprenants (Greenhow and Askari 2017) et sont considérés comme un puissant moteur de changement pour les pratiques d'apprentissage, en termes d'ouverture, d'interactivité et de sociabilité (Manca and Ranieri 2016).

Après avoir examiné les facteurs d'abandon dans les MOOC à l'aide de la littérature et souligné le potentiel de l'utilisation des informations des médias sociaux et de leurs caractéristiques pour promouvoir l'engagement et la satisfaction de l'apprenant, nous visons à résoudre le problème de l'abandon en faisant coopérer les informations des deux environnements (MOOC et médias sociaux). L'objectif de notre thèse est d'exploiter les informations et l'implication personnelle de l'apprenant dans les médias sociaux dans le but d'enrichir et de personnaliser son profil dans les MOOC. Comme nous le savons, les apprenants des MOOC ont des comportements d'apprentissage, des habitudes d'apprentissage et des temps d'apprentissage très différents, ce qui conduit à des profils d'apprenants et des besoins d'apprentissage différents.

2. Contributions de la thèse:

Notre travail met en évidence le potentiel des ontologies et du NLP pour améliorer l'expérience d'apprentissage en ligne et pour répondre aux besoins et aux préférences des profils individuels des apprenants. Plus précisément, nous avons mis en œuvre des ontologies pour créer le profil

social de l'apprenant et un algorithme de modélisation des sujets pour détecter les cours d'intérêt des apprenants à partir de leur contenu généré sur les médias sociaux (tweets).
Pour être plus précis, notre approche a fait l'objet de plusieurs étapes de validation, comme suit:
Premièrement, nous avons étudié les défis qui entourent les MOOC et les facteurs à l'origine de l'abandon, nous avons découvert qu‘au moins d'interaction et de sociabilité, le manque d'engagement continu et de motivation et les caractéristiques des apprenants sont quelques-unes des raisons qui ont contribué à l'arrêt du processus d'apprentissage.
Deuxièmement, afin de refléter l'utilité des médias sociaux dans l'apprentissage en ligne, nous avons utilisé un cadre d'analyse appelé analyse SWOT, pour examiner les facteurs internes et externes, ainsi que le potentiel actuel et futur des MOOC et des médias sociaux. Les résultats de cette analyse ont montré que les deux environnements pouvaient avoir des fonctions complémentaires (Zankadi et al. 2018).
Troisièmement, nous nous sommes concentrés sur les caractéristiques et les interactions des apprenants dans les MOOC et les médias sociaux. Nous avons d'abord interrogé les éléments qui constituent un profil d'apprenant, pour mieux comprendre et mettre en évidence les caractéristiques et les besoins de l'apprenant dans les MOOC et les médias sociaux (Zankadi et al.2018). Ensuite, nous avons utilisé l'ontologie pour créer le profil social de l'apprenant (Zankadi et al. 2022).
Les ontologies fournissent une description formelle des définitions des classes conceptuelles et de leurs relations qui va au-delà des listes, des thésaurus et des taxonomies. Par formel, nous entendons que les définitions sont fondées sur un cadre logique (Harrow et al. 2019). Les ontologies sont utilisées comme méthode d'encodage de la sémantique d'un domaine de connaissance humaine sous une forme lisible par une machine.
L'ontologie a été décrite pour la première fois par Gruber comme une "spécification explicite d'une conceptualisation" en 1993, (Staab et Studer 2009). Dans ce contexte et selon (Al-Yahya, George et Alfaries 2015):

- Une conceptualisation est une représentation abstraite d'une certaine partie du monde qui prend la forme d'une définition des caractéristiques des termes clés.
- Une spécification explicite signifie que le modèle doit donner un sens au vocabulaire tout en le rendant exploitable par les machines et les humains.

Le processus de génération de l'ontologie globale qui représente notre profil social s'est déroulé en trois grandes étapes:

Étape 1: Processus de création d'ontologies locales en utilisant la méthode METHONTOLOGY

La METHONTOLOGIE, l'une des méthodologies les plus connues et les plus complètes pour le développement d'ontologies, sert de base à l'approche méthodologique utilisée pour construire, mettre en œuvre et représenter l'ontologie dans notre travail. La METHONTOLOGIE est une approche bien structurée qui consiste en un certain nombre d'activités, de techniques pour réaliser chaque activité et de produits livrables à créer après la réalisation des activités à l'aide des techniques (Fernández-López, Gómez-Pérez, and Juristo 1997). Parmi les tâches réalisées au cours de cette phase de construction des ontologies locales:

1. **Spécification de l'ontologie:**

- **Objectif:** élaborer un document de spécification des exigences relatives à l'ontologie (ORSD). Il décrit l'objectif principal, la fonction, le niveau de granularité et la portée de l'ontologie.
- Domaine: MOOCs et médias sociaux.
- Objectif : Construire un profil formel d'apprenant social intégrant des informations sur les MOOC et les médias sociaux.
- Portée: Concepts incluant les intérêts de l'apprenant, ses préférences, son comportement, ses styles d'apprentissage, etc.
- Exigences : Dérivé des défis d'apprentissage de la vie réelle.

2. **Acquisition de connaissances :**

- **Objectif :** recueillir les connaissances du domaine pour la création d'une ontologie.
- Approche de l'analyse documentaire pour la modélisation d'ontologies dans les MOOC et l'étude des profils d'utilisateurs dans les médias sociaux.
- Connaissances tirées de livres, d'articles, de normes et d'ontologies existantes.
- Définition des concepts, des relations et des propriétés pour les classes primaires de l'ontologie.

3. **Conceptualisation :**

- **Objectif:** organiser les connaissances du domaine dans un modèle conceptuel.
- Les tâches comprennent la création d'un glossaire de termes, de taxonomies de concepts, de relations binaires ad hoc, d'un dictionnaire de concepts, etc.
- Utilisation de relations taxonomiques telles que Subclass Of, Disjoint-Decomposition, Exhaustive-Decomposition et Partition.
- Les diagrammes de relations binaires ad hoc et les dictionnaires de concepts détaillés aident à structurer les connaissances.

4. **Intégration:**

- Identification des termes d'autres ontologies à inclure.
- Identification de nos attributs personnalisés à inclure dans l'ontologie de profil social.
- Utilisation de normes telles que FOAF, IMS LIP, la taxonomie de Felder pour l'ontologie MOOC.
- SIOC et l'ontologie Emoji pour les profils d'utilisateurs des médias sociaux.

5. **Mise en œuvre:**

- Utilisation de Protégé v5 pour la formalisation de l'ontologie.
- Mise en œuvre d'ontologies distinctes pour les profils d'apprenants dans les MOOC et les médias sociaux.

6. **Évaluation:**

- Évaluation basée sur des critères : Clarté, cohérence, concision, exactitude.
- La contribution d'experts et l'examen fondé sur des critères ont permis d'améliorer l'ontologie.
- Prise en compte des commentaires sur la clarté, la cohérence, la concision et l'exactitude de l'ontologie.

Étape 2: Processus de mise en correspondance des ontologies à l'aide de COMA 3.0

Étant donné deux ontologies A et B, la mise en correspondance d'une ontologie avec une autre signifie que pour chaque concept (nœud) dans l'ontologie A, nous essayons de trouver un concept (nœud) correspondant, qui a la même sémantique ou une sémantique similaire, dans l'ontologie B et vice versa" (Ehrig and Sure 2004).
L'outil d'algorithme d'appariement COMA 3.0 met en œuvre une approche itérative basée sur une variété d'algorithmes d'appariement (matchers) pour la mise en correspondance d'ontologies (Li et al. 2016). Nous avons utilisé trois algorithmes d'appariement individuels : "AllContextW", "NodesPathW" et "NodesNamesW". Des mesures d'évaluation telles que Recall, Precision et F-measure ont été utilisées pour évaluer la qualité de l'apparieur. Sur la base de l'évaluation, l'outil de mise en correspondance "NodesPathW" a été choisi pour la mise en correspondance d'ontologies en raison de sa précision supérieure.

Étape 3: Fusion d'ontologies à l'aide de COMA 3.0

La fusion d'ontologies est un processus qui consiste à combiner deux ou plusieurs ontologies en une seule. Par conséquent, l'ontologie résultante stocke les connaissances de toutes les ontologies fusionnées. La fusion utilise souvent un ensemble d'alignements pour créer des interconnexions profondes entre les ontologies et, à la fin, les fusionner en une seule.
Nous avons également utilisé COMA 3.0 pour le processus de fusion. Nous testons les différentes combinaisons basées sur les résultats des correspondances afin d'obtenir le meilleur résultat de fusion. Sur la base de l'évaluation des outils d'appariement, nous avons fusionné les résultats de l'outil d'appariement "NodesPathW" et de l'outil d'appariement "NodesNameW".
Le processus de fusion des ontologies combine les ontologies mises en correspondance pour créer l'ontologie globale (SPOnt), qui représente le profil social intégré de l'apprenant. Cette ontologie peut servir d'outil pour offrir des possibilités d'apprentissage personnalisées aux apprenants en identifiant leurs préférences et leurs intérêts et en les guidant vers les cours et les ressources appropriés.
Quatrièmement, sur la base de l'ontologie des profils sociaux, nous nous sommes concentrés sur la classe "Intérêt" comme étant une partie importante d'un profil d'apprenant. Dans le contexte des MOOC, les intérêts jouent un rôle essentiel dans le processus d'apprentissage, ils constituent un processus de motivation puissant qui dynamise l'apprentissage et guide les trajectoires académiques et professionnelles (McIntyre, Gundlach, et Graziano 2021) (Harackiewicz, Smith, et Priniski 2016). Les apprenants sont motivés pour investir du temps et des efforts dans les cours qui les intéressent. Ainsi, l'enrichissement de l'intérêt des apprenants permettra une meilleure découverte des sujets de cours qui correspondent le mieux à leurs préférences, ce qui aura un impact sur leur satisfaction et donc sur leur interaction au sein des MOOC.
Les MOOC recommandent déjà des cours qui répondent aux intérêts des apprenants en fonction de leur participation. Cependant, ces mêmes apprenants sont plus interactifs dans les médias sociaux grâce au contenu qu'ils génèrent et qui contient des informations cachées sur leurs intérêts et préférences "réels". Étant donné que générer l'intérêt de l'utilisateur est une tâche difficile, l'utilisation de techniques de modélisation des sujets est utile pour découvrir les principales informations thématiques liées à l'utilisateur (Bai et al. 2021).
Nous proposons un modèle de sujet de cours (CTM) basé sur le traitement du langage naturel (NLP) et des techniques de modélisation de sujet pour identifier et détecter le cours d'intérêt des apprenants sur la base de leur interaction spontanée dans les médias sociaux, en particulier

Twitter. Le CTM généré contient le sujet d'intérêt le plus probable pour chaque apprenant (Zankadi, Idrissi, et al. 2022).
L'approche proposée comporte plusieurs étapes séquentielles pour la génération de sujets. Ces étapes comprennent la préparation des données, le prétraitement des données et l'extraction des caractéristiques. Trois méthodes de modélisation des sujets - LDA, LSA et BERTopic - sont entraînées et la performance du modèle est évaluée à l'aide du score de cohérence des sujets, du score de diversité des sujets et du jugement humain :

Étape 1: Collecte et prétraitement des données

1. **Collecte des données:**
   - Tweets collectés par web scraping à l'aide des clés API de Twitter et de Netlytic.
   - Filtres appliqués pour la langue anglaise et les mots-clés "Computer science" et "Artificial intelligence".
   - Ensemble de données: 120 000 tweets de 12 187 apprenants avec métadonnées.
2. **Prétraitement des données:**
   - Les doublons et les utilisateurs de chatbot ont été supprimés à l'aide de la reconnaissance des chaînes de caractères.
   - Tweets nettoyés en supprimant les liens, les chiffres, les emojis et la ponctuation.
   - Bibliothèque tweet-preprocessor utilisée pour les étapes de prétraitement (URL, Mentions, Emojis, Nombres, Hashtags, espaces blancs et ponctuation).

Étape 2: Tâches NLP

1. **Tokenisation:**

Les phrases sont divisées en mots, en minuscules, les mots courts sont ignorés et les accents de lettres sont supprimés.

2. **Suppression des mots vides:**

Suppression des mots vides ne contenant pas de contenu ; la liste est complétée par les mots à haute fréquence.

3. **Mise en œuvre de N-grammes:**

Les bi-grammes et les tri-grammes sont implémentés à l'aide du modèle Gensim Phraser.

4. **POS/Speech of Tag Selection:**

Marquage des parties du discours à l'aide de la bibliothèque Gensim ; conservation des noms, des adjectifs, des verbes et des adverbes.

5. **Lemmatisation:**

Les terminaisons flexionnelles sont supprimées et la forme de base des mots est conservée.

6. **Extraction des caractéristiques du texte:**
   - Bag of Words (BoW) et TF-IDF appliqués pour créer des corpus.
   - Filtrage des mots à haute fréquence, création d'un dictionnaire et génération de corpus BOW et TFIDF.

Étape 3: Paramètres et formation du modèle

1. **Formation LDA et LSI:**
   - Choix du nombre de sujets latents ; recherche du nombre optimal sur une grille.
   - A priori de Dirichlet réglés sur "Auto".
2. **Formation BERTopic:**
   - Transformateur Countvectorizer utilisé pour le prétraitement.

- BERTweet utilisé comme modèle de transformateur de phrases.

Étape 4: Évaluation et présentation des sujets

1. **Évaluation des sujets:**

Le score de cohérence, le score RBO (Inverted Rank-Biased Overlap) et le jugement humain sont utilisés pour comparer les modèles.

2. **Représentation des thèmes à l'aide d'un nuage de mots:**
   - Visualisation du nuage de mots pour représenter les sujets.
   - Étiquetage manuel des sujets pour les modèles LDA et BERTopic.

Conclusion:

Cette thèse explore l'intersection des MOOC et des médias sociaux en utilisant des approches basées sur l'IA. Notre recherche applique les ontologies et le NLP pour améliorer l'expérience d'apprentissage en ligne afin de répondre aux besoins et aux préférences des profils individuels des apprenants. Plus précisément, nous avons mis en œuvre une ontologie de profil d'apprenant social afin de créer une vue plus complète de l'apprenant.

Notre objectif est d'intégrer des informations provenant de différentes sources (MOOC et médias sociaux) et de les combiner pour obtenir une compréhension plus complète des intérêts, des capacités et des besoins de chaque apprenant.

À cette fin, nous avons créé deux ontologies locales pour représenter le profil de l'apprenant dans le MOOC et le profil de l'utilisateur dans les médias sociaux. Sur la base de normes et d'ontologies bien connues telles que IMS LIP, FOAF, l'ontologie des organisations, l'ontologie SIOC, etc., nous avons proposé nos propres attributs pour créer un profil plus personnalisé. Ensuite, nous nous appuyons sur des techniques d'ingénierie ontologique telles que le mappage et la fusion d'ontologies pour mettre en œuvre notre ontologie de profil social.

En utilisant l'ontologie précédente, nous nous sommes concentrés sur la composante "Intérêt", pour détecter les sujets d'intérêt potentiels des apprenants à partir de leur contenu sur les médias sociaux (tweets). Nous avons mis en œuvre un pipeline NLP et appliqué des algorithmes de modélisation thématique tels que LDA, LSA et BERTopic sur les tweets. Cela pourrait fournir des informations sur les ressources éducatives qui correspondent mieux au profil de l'apprenant.

Notre thèse est un travail en cours qui a donné des résultats prometteurs, y compris notre engagement à partager l'ontologie avec la communauté, à assurer sa "réutilisabilité" et à fournir une documentation complète pour l'accessibilité.

En conséquence, il est intéressant de poursuivre certaines perspectives comme suit:

- Garantir une éducation inclusive et des possibilités d'apprentissage tout au long de la vie en tenant compte d'une plus grande personnalisation du profil des apprenants ayant des besoins particuliers, tels que les personnes souffrant d'une déficience auditive ou visuelle et les personnes souffrant d'un trouble du déficit de l'attention/hyperactivité (TDAH) et d'un trouble du spectre autistique (TSA).
- Utilisation d'une plateforme MOOC réelle, telle que EDX ou Coursera, pour améliorer leur système de recommandation avec les résultats de notre approche, afin de quantifier son impact sur le taux d'abandon, l'engagement des apprenants et l'achèvement des cours.

- Explorer les différences de comportement vis-à-vis des pratiques d'apprentissage, concernant les différentes générations telles que les baby-boomers, la génération X, les milléniaux, la génération Z et la génération Alpha. La découverte de ces différenciations particulières pourrait enrichir et étendre le profil personnalisé proposé par notre approche afin de mieux répondre à leurs attentes.
- En outre, la révolution actuelle alimentée par de grands modèles linguistiques tels que GPT-3.5 pourrait améliorer l'assistance à l'apprenant et favoriser une expérience d'apprentissage plus interactive et dynamique. Celle-ci pourrait être enrichie par l'intégration de ChatGPT dans les plateformes MOOC, afin d'engager des conversations personnalisées basées sur le profil social fourni par notre approche. Par conséquent, notre approche renforcera ses capacités de traitement du langage naturel, ce qui permettra d'apporter des réponses plus personnalisées et contextuelles aux apprenants des MOOC.
- Enfin, le potentiel de l'incorporation de la réalité virtuelle (RV) combinée à notre approche pourrait être étudié pour améliorer l'engagement de l'apprenant et la rétention des connaissance.

List of Figures

List of Tables

List of Abbreviation

Abbreviation	Meaning
AI	Artificial Intelligence
BOW	Bag of Words
CF	Collaborative Filtering
cMOOC	Connectivist Massive Open Online Course
CMS	Contents Management System
COVID	Coronavirus Disease
cTF-IDF	Class Term Frequency Inverse Document Frequency
CTM	Course Topic Model
DAML	DARPA Agent Markup Language
FL	Fuzzy Logic
FOAF	Friend Of A Friend
FSLSM	Felder-Silverman Learning Style Model
ICT	Information and Communication Technologies
IMS LIP	IMS Learner Information Package
LCMS	Learning Contents Management System
LDA	Latent Dirichlet Allocation
LMS	Learning Management System
LSA	Latent Semantic Analysis
LSTM	Long Short Term Memory
MOOCs	Massive Open Online Courses
Moodle	Modular Object-oriented Dynamic Learning Environment
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
OIL	Ontology Inference Layer
ORSD	Ontology Requirement Specification Document
OWL	Web Ontology Language
QCL	Qualifications, Certifications, Licenses
RDF	Resource Description Framework
RDFs	Resource Description Framework Schema
SIOC	Semantically-Interlinked Online Communities
SPOnt	Social Profile Ontology
SRWL	Semantic Web Rule Language
SVD	Singular Value Decomposition
SWOT	Strengths, Weaknesses, Opportunities and Threats
TF-IDF	Term Frequency Inverse Document Frequency
TM	Topic Modeling
WSD	Word Sense Disambiguation
xMOOC	Instructivist MOOC

Table of Contents

General Introduction

E-learning is a topic of discussion since 1990s but its evolution has attracted considerable attention especially during the COVID 19 pandemic, where switching to online learning is now required. E-learning is defined as *''The acquisition of knowledge through electronic technologies and media* (S.-H. Kim and Park 2021) (Milićević et al. 2021)'' in an effort to provide learning resources that are more widely accessible, free from the constraints of space, time, and distance, to learners all over the world. This has made e-learning one of the most well-liked sectors of the global economy and one of the most widely used educational approaches in the 21st century.

Being a highly efficient, adaptable, and cutting-edge e-learning tool, MOOCs have attracted attention. Their remarkable benefits of openness, limitless access to materials, curriculum sharing, and low cost have contributed to their rapid growth. They excelled in disciplines related to the humanities, business management, health, and sciences, which increase enrollments worldwide and learners' engagement via dedicated web portals, irrespective of location. They have shown to be an effective teaching and learning tool and an alternative to conventional face-to-face learning techniques. Low course completion rates and high dropout rates, however, have a significant impact on MOOC participants' perseverance.

Many factors influence the learners' decisions in all stages of MOOCs acceptance, including feeling of isolation and less interaction, lack of continuous motivation, learner characteristics, MOOCs features and learner experience.

On the other hand, there is a rising reliance on social media due to its broad use (e.g., Facebook, Twitter, YouTube, Instagram, TikTok, etc.). Its connectivity, interactivity, and user-generated content are its defining features. Furthermore, social media has a significant impact on information sharing through interactive communication that reflects users' thoughts, sentiments, and actions. As a result, the use of social media and its effects on socializing provide a reliable source for gathering and analyzing users' preferences and interests.

In this context, social media-supported e-learning can enhance engagement and cooperation by bringing people together who have similar goals and interests, promoting active participation, increasing the variety and accessibility of learning, and unifying and sharing content. Therefore, including social media information and incorporating its features into the teaching and learning processes in MOOCs can have a positive impact on the learners. According to (Salloum, Al-Emran, et al. 2021), social media aspects have modernized e-learning by introducing new functionalities like reacting, commenting, motivating, or group creation, they have a major beneficial impact on perceived usefulness and perceived ease of use of e-learning environments. Giving learners social e-learning that supports their interests and preferences can thereby boost their role in sociability, feeling of community, and course satisfaction and thereby encourage engagement.

The main objective of our thesis is to contribute to the field of e-learning by gaining a deeper understanding of learners needs and their learning requirements. We aim to provide learners with a more personalized and individualized learning experience within MOOC by enriching his profile using his information from social media profile. Each learner is different from other in term of interests, preferences, learning styles, etc. So, building and implementing a social profile that captures the knowledge about the learners will lead to a personalized learning content that meets each learner requirements and therefore impact their engagement and interaction inside MOOCs. Our approach is based on leveraging social media for social and personalized MOOCs through learner profiling and course interest detection using ontologies and NLP methods such as topic modeling algorithms.

Thesis contribution:

The main aim of our thesis is to better enhance and personalize the learner experience within MOOC by enriching his profile using his information from social media profile. Our research approach explores the intersection of MOOCs and social media using AI-driven tools.

Our work highlights the potential of ontologies and NLP to improve the e-learning experience and to meet the needs and preferences of individual learners' profiles. Specifically, we have implemented ontologies to create the social learner profile and a topic modeling algorithm to detect learners' course of interest from their generated content on social media (tweets).

To be more specific, our approach encountered several validation steps as the following:

First, we investigated the challenges that surround MOOCs and the factors behind the drop-out, we found out that: less interaction and sociability, lack of continuous engagement and motivation and learners' characteristics are some of the reasons that contributed to stopping the learning process.

Second, in order to reflect the usefulness of social media in e-learning, we used an analysis framework called SWOT analysis, to examine the internal and external factors, as well as current and future potential of both MOOCs and social media. The output of this analysis showed that both environments could have complementary functions (Zankadi et al. 2018).

Third, we focused on learners 'characteristics and learners 'interaction inside MOOC and social media. We interrogated first the elements that constitutes a user profile, to better understand and highlight both the characteristics and the requirement of the learner inside MOOCs and social media (Zankadi et al.2018). Then, we used ontology to create the learner social profile (Zankadi, Hilal, et al. 2022). The process of generating the global ontology that represents our social profile went through 3 major steps:

- Creating two local ontologies: Local ontology 1 that represents the profile of the learner within MOOC including learning style, preferences and other personal information and Local ontology 2 that represents the user profile in social media and which include interests, preferences, posts, etc.;
- Ontology mapping: we mapped the two local ontologies;
- Ontology merging: to merge the mapped ontology

Forth, based on the social profile ontology, we focused more on the 'Interest' class as being an important part of a user profile. We identified and detected the course of interest of learners from their generated content in social media (Twitter), by applying topic modeling and NLP techniques on the textual feature. We implemented many unsupervised machine learning methods such as Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA) and BERTopic after applying NLP pipeline on the tweets shared by the learners. The evaluation of the models reveals that BERTopic model performed better on the scrapped dataset and their results are used to generate the course topic model. The extracted topics represents the potential course of interest of the learners (Zankadi, Idrissi, et al. 2022).

The thesis structure:

The thesis is organized into five chapters. It is structured as follows.

Chapter 1 highlights the social e-learning and learner profiling background. It points out the different definitions, concepts and techniques used during the implementation of the ontology-based social profile.
Chapter 2 presents an overview of the literature review related to the use of social media for e-learning, the different works that highlighted the learner profiling as well as works that focused on using topic modeling within e-learning platforms.
Chapter 3 outlines our proposed contribution related to the implementation of the ontology-based social profile. It presents the methodology adopted for the implementation of the ontology, as well as the different techniques used to create the social profile ontology.
Chapter 4 introduces our course topic model to identify and extract the topical interest from the text content shared by learners on social media to enrich their course preferences in MOOCs. We apply NLP pipeline and topic modeling techniques to the textual feature using three well-known topic models: Latent Dirichlet Allocation, Latent Semantic Analysis, and BERTopic.
Chapter 5 presents an overview of the application of recommender systems in e-learning. We point out as well the different works that highlight the use of recommender systems in e-learning using literature review.
Finally, the thesis is concluded by summarizing the major outcomes of this thesis, the list of our publications and perspectives that can be explored and points to the directions in which this research can be continued.

Chapter 1: Context and background of social e-learning and learner profiling

Chapter 1 Context and background of social e-learning and learner profiling

Introduction:

In this chapter, we delve into the evolution of e-learning, exploring the shift from its early stages to modern prominence. Tracing the trajectory from early Computer-Based Training (CBT) experiments to the contemporary dominance of Massive Open Online Courses (MOOCs), we examined not only technological advancements but also the changing paradigms of teaching and learning. Central to this evolution is Bandura's social learning theory, which posits that learning occurs in a social context through observation and modeling. This theory provides a foundation for understanding how learners internalize and interpret observed behaviors.

The discussion expanded to the realm of social e-learning environments, emphasizing the significance of knowledge sharing in social and cultural contexts. Social media, particularly in the context of e-learning, emerged as a positive influencer, fostering engagement, interaction, and improved self-efficacy among learners. Learners actively participate through various forms of digital interaction, contributing to a rich learning experience.

The latter part of the chapter focused on user profile modeling, elucidating the concept of user profiles and their role in understanding learners' cognitive skills, preferences, and behavior. User profiling involves diverse techniques, including the overlay model, stereotype model, vector representation, and fuzzy logic modeling. Furthermore, the chapter introduced ontology-based user profiling, highlighting the use of ontologies to represent user interests and enhance personalized learning experiences.

1. Evolution of the concept E-learning:

1.1. E-learning evolution:

The definition of e-learning has evolved throughout time. Table 1 presents the different definitions attributed to the term ‘e-learning’ between 2000 and 2021 as presented in (Choudhury and Pattnaik 2020).

Table 1 Different definitions attributed to the term 'e-learning' between 2000 and 2021

Year	Definition	Authors
2000	The acquisition and use of knowledge distributed and facilitated primarily by electronic means	(Wentling et al. 2000)
2001	The use of internet technology to supply a wide range of learning solutions and provide the correct information to the right person at the right time, as well as to expand the person's knowledge base and change how they perceive the actual world	(De Cubber 2001) (Govindasamy 2001)
2002	learning using internet-based electronic device	(Rosenberg and Foshay 2002); (Barker 2002)
2003	Any educational or training endeavor that relies on information and communication technologies to enhance learning activities whenever a learner gathers knowledge	(Sambrook 2003)

	electronically without an instructor being physically present there.	
2005	A distributed, web-based learning environment that makes use of educational technologies to encourage learning and knowledge acquisition through significant activity and interaction	(Gilbert and Dabbagh 2005)
2006	The adoption of any new technologies or programs to support learning or learner assistance	(Laurillard 2005)
2010	Electronic technologies, such as the Internet, Web 2.0, intranets, and extranets, that distribute or facilitate educational content or learning experiences	(Njenga and Fourie 2010); (Bondarouk and Ruël 2010); (Klobas and McGill 2010)
2012	E-learning is defined as follows: (1) Technology-driven: utilizing technology to deliver learning and training programs; (2) Delivery-system-oriented: utilizing electronic means to deliver learning, training, or education programs; (3) Communication-oriented: Learning made possible by the use of digital tools and materials that encourage interaction, such as online communication between students and their teachers or peers; (4) Educational-paradigm-oriented: Information and communication technologies used to help students improve their learning.	(Sangrà, Vlachopoulos, and Cabrera 2012)
2013	The application of ICT to make information and knowledge resources available to learners while removing time and location constraints. The use of the Internet and new multimedia technologies to improve the quality of learning by granting access to resources and services as well as facilitating remote exchange and cooperation is known as e-learning, according to the European Commission.	(Dominici and Palumbo 2013); (Allen and Seaman 2013); (Upadhyaya and Mallik 2013)
2014	E-learning is a type of distant education that uses an electronic channel (medium), such as the Internet, to entirely virtualize the learning experience.	(Lara et al. 2014)
2015	The acquisition of knowledge and skills via the use of information and communication technologies (ICTs), with a focus on fostering interactions between learners and tools, material, and learning activities.	(Tîrziu and Vrabie 2015)

2016	Using modern tools to access educational resources while learning outside of a regular classroom. E-learning pedagogies are categorized into four generation: behaviorist/cognitivist, social constructivist, connectivist, and holistic. Each pedagogical strategy's applicability depends on the technological tools it employs.	(Arun Gaikwad 2016); (Dron and Anderson 2016); (Kattoua and Al-Lozi 2016)
2017	A revolutionary innovation in educational technology that is currently altering how learning is approached.	(Garrison 2016)
2019	A cutting-edge web-based system built on technological innovations and various types of educational materials, with the main objective of giving learners a personalized, learner-centered, open, engaging, and interactive learning environment that supports and enhances the learning processes.	(Rodrigues et al. 2019); (Almaiah and Al Mulhem 2019)
2020	▪ The process of using technology to learn. ▪ It is regarded as a set of instructions that are sent through all electronic media, including the internet, intranets, and extranets. learners can control their own lifelong learning by removing the constraints of time and geography.	(Al-Fraihat, Joy, and Sinclair 2020); (Al Mulhem 2020)
2021	▪ e-learning defined as an information system that can integrate a variety of educational materials via email, discussion; assignments; quizzes; and live chat sessions.	(Suzianti and Paramadini 2021); (Amarneh et al. 2021); (Rawashdeh et al. 2021)

Defining e-learning proves challenging due to ongoing technological advancements. It is described as a technological learning process encompassing instructions transmitted through electronic media. E-learning allows learners to self-regulate lifelong learning, overcoming time and geographical constraints, while some characterize it as an information system integrating various educational materials. Based on criteria such as networking, device delivery, and a broad learning perspective, e-learning offers real-time updates, information storage, distribution, and a departure from conventional educational paradigms.In the next section, we will explore the existing types of e-learning environments.

1.2. E-learning environments :

The Learning Management System (LMS), Course Management System (CMS), Learning Content Management System (LCMS) are a few of these software programs. Each of them offers a particular kind of assistance and features to the educational organization or institutions that deployed them.

1.2.1. LMS:

Learning Management System (LMS) is a web-based system that aids in planning, disrupting, and analyzing any learning process with a focus on managing learners, keeping track of their performance, and tracking their progress with the activities. It is an environment software program created for controlling learner interactions and supplying them with educational materials (Alshammari, Bilal Ali, and M.S. 2018) (Jung and Huh 2019; S. Kim and Huh 2019). There are many LMS examples, but the most well-known one is Modular Object-oriented Dynamic Learning Environment (Moodle), an open-source program created in 2002. It was created in accordance with the GNU license, which stands for General Public License (GBL), which permits any modification to the source code as long as the license of the original source is preserved (Ekwonwune and Edebatu 2019).

1.2.2. CMS:

A CMS (Contents Management System) is any program that makes it easier to create and publish digital information. Platforms for generating blogs, forums, online stores, static sites, and everything in between fall under this category. In addition to these well-known CMSs, there are also Joomla, Drupal, Shopify, and Squarespace, all of which have a substantially smaller market share (Cabot 2018; S. Kim et Huh 2019), (Cabot 2018; S. Kim and Huh 2019).

1.2.3. LCMS:

The Learning Contents Management System, or LCMS for short, is a learning object management system that offers the ability to manage learning materials installed in LMS. A learning target may be mounted, modified, or removed as part of management. Also, it has a function that allows you to search for preferred learning materials (Jung and Huh 2019).

1.3. MOOCs environments:

1.3.1. MOOCs:

Massive Open Online Courses (MOOCs), often known as "Connectivist" learning models, were developed by George Siemens and Stephen Downes (Downes 2005). Massive Open Online Courses (MOOCs) are cutting-edge online learning resources that are accessible to a wide audience without regard to time or geography (Wong et al. 2019). They are web-based courses designed and delivered by approved higher education institutions and organizations in which anyone with a smart device and internet connection can participate, regardless of age, gender, geographic location, or education background (Deng, Benckendorff, and Gannaway 2019).
Openness, participation, and distribution are the three main qualities of a MOOC (Baturay 2015) (Al-Imarah and Shields 2019) (D. Tao et al. 2022):

- **Open:** Anyone with Internet access is welcome to sign up for a MOOC at no cost.
- **Participatory:** In a MOOC, learning is improved by participation in both the development and sharing of individual contributions as well as interactions with those of others. Yet, participation is optional and has no impact on the educational system.
- **Distributed:** Because MOOCs follow the connectivist approach, all knowledge should be shared among the network of participants.

1.3.1.1.cMOOCs:

The first MOOCs denoted as "cMOOCs" (Connectivism MOOCs) first debuted in 2008 and were based on the "connectivist distributed peer learning paradigm," which refers to the connection of learners through electronic networks and their ability to participate in collaborative learning.
According to (García-Peñalvo, Fidalgo-Blanco, and Sein-Echaluce 2018), the structural model of a cMOOC is depicted in Figure 1, where each participant (P) creates resources (R) that are distributed to the other participants. This structure is typical of the Web 2.0 environment and it is perfectly suited to the new networked learning model, and offers a disruptive opportunity for the traditional university learning paradigm.

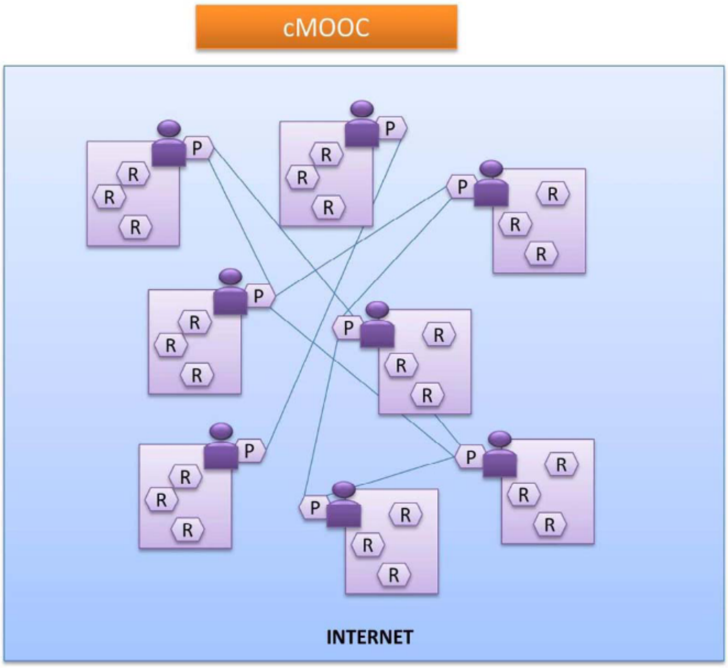

Figure 1 cMOOC structure model

xMOOCs (content-based MOOCs) was introduced in 2011. They are characterised by being massive and self-directed (Flores et al. 2020). Private learning management systems, as highlighted by (García-Peñalvo, Fidalgo-Blanco, and Sein-Echaluce 2018), play a crucial role in delivering xMOOCs, emphasizing structured content and assessments in online education. As in figure 2, the xMOOC, teachers produce and structure resources and activities in the platform, where learners access the materials.(García-Peñalvo, Fidalgo-Blanco, and Sein-Echaluce 2018).

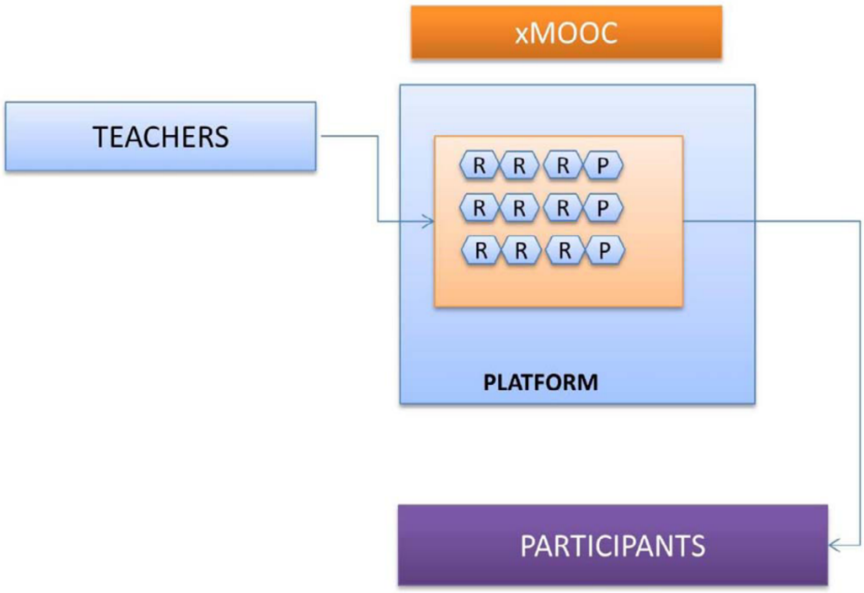

Figure 2 xMOOC structure model

Table 2 presents the main differences between cMOOC and xMOOC, according to (Smith and Eng 2013), (Saadatdoost et al. 2015) and (Mohamed and Hammond 2018):

Table 2 The key differences between cMOOC and xMOOC

Features	cMOOC	xMOOC
Course Content	▪ The course is not defined by its content. ▪ Being exploratory in style. ▪ As a starting point, certain course materials are made available online. ▪ A forum for interaction and knowledge formation among students through their joint contributions.	▪ The content is focused and packaged. ▪ Controlled. ▪ Comparable to the conventional in-class method. ▪ Online access to all course materials. ▪ Video presentations and reading text are predominant.
Interaction	▪ Predominantly peer to peer ▪ Monitored by instructors	▪ Instructor feedback ▪ Peer feedback

	▪ Inspired to take part in forums decentralization and social media	▪ Most discussion forums are located on the course website.
Assessment Methods	▪ Formative ▪ Summative ▪ Instructor-graded ▪ Peer-reviewed assessments	▪ Formative ▪ Summative ▪ Automated ▪ Peer-reviewed assessments
Instructor/Learner Roles	▪ Non-traditional: "distributed, chaotic, emergent" ▪ Learners are required to develop, expand their scope of knowledge, and share their own sense-making through producing artifacts. ▪ Self-directed learners	▪ Conventional: imparting knowledge to learners ▪ Learners' acquisition of knowledge and mastery
Pedagogy	Connectivist	Cognitive-Behaviorist

Table 2 presents a comparison of the key characteristics of both cMOOC and xMOOC. They differ primarily in how much knowledge is duplicated or created, which is correlated with openness.

In cMOOC, learners are encouraged to create the content for their own learning through cooperative networking. Due to this, cMOOCs are developed without a centralized core content. Therefore, the opportunities available through cMOOCs are generally wide-open. In short, cMOOCs emphasize the values of autonomy and variety.

However, with xMOOCs, openness is either less common or defined differently. They are centralized networks with a single major platform, and designers often determine the course content. Discussion boards are where interactions occur the majority of the time.

Furthermore, a key factor in choosing these platforms is the difference in community longevity between cMOOC communities and xMOOC communities. As a result, xMOOCs encourage learners to use social networking sites to overcome the course's short duration's restriction and continue building relationships after it is ended.

After introducing the evolution of e-learning definitions and presenting e-learning environments and MOOCs environments, we focus in the next section, on social e-learning, a new form of learning integrating social media aspects.

2. Social e-learning:

Bandura explained learning as the process through which a person gains new knowledge by observing others and imitating their behavior. The process is called social learning, as observation and modeling occur in social contexts once learners mimic others' actions, whether they do so intentionally or unknowingly. (Rumjaun and Narod 2020) (Ahn, Hu, and Vega 2020). According to (Pahl-Wostl and Hare 2004), "Social learning is predicated on iterative feedback between the learner and their environment, as well as the learner altering their environment and these changes influencing them.”.

Bandura’s theory of social learning offers a helpful foundation for thinking about how learners learn through modeling and observation (Bandura and Walters 1977). According to Bandura, learning occurs in a social context through observation, but it also involves cognitive processes; in other words, learners internalize and interpret what they observe in order to replicate the behavior themselves (Horsburgh and Ippolito 2018).

The social learning theory connects learning to its social context and perspectives. knowledge must be shared among learners in social and cultural contexts, such as social learning environments, for social learning to take place. Knowledge is defined as content plus the socially constructed value in these environments, which are created by learners utilizing a variety of social technologies, such as social networking, tagging, blogging, etc. Hence, social learning implies an ongoing multi-actor negotiation process between learners and the setting in which they are involved, ultimately leading to knowledge co-creation (Deaton 2015).

Recently, the use of social media in the context of e-learning has impacted positively the learning outcomes. Learners are actively engaged and interact through diverse form like, posting a comment, reading an article, liking a post, or retweeting a message. Social learning theory-recommended interactions are promoted in the social media environment, and as digital interactions are often free of many social anxieties, users frequently show a higher level of self-efficacy with regard to the experience. This increased self-efficacy could lead to greater engagement, which in turn may result in improved learner learning. Via a variety of user-generated material, social networking enables users to communicate and work together fully (e.g., videos, images, audio, text posts). Users can therefore emulate actions in movies or images created by or shared by other users in their network (Peng et al., 2019). Users who are learning can learn underlying cultural norms from text posts that articulate other users' problem-solving strategies (Deaton, 2015) (Castellanos-Reyes, Maeda, and Richardson 2021).

In social e-learning environments, learners engage in collaborative activities, share insights, and contribute to the creation of a collective knowledge space. In order to capture these social interactions, user profile modeling comes into play to shaping personalized learning experiences based on individuals' contributions and preferences within the social context of e-learning platforms.

3. User profile modeling:

3.1.User profiling modeling:

A user profile describes a sample of the knowledge related to the user and to its context. It highlights an insight about both direct and indirect user’s information in a structured way. According to (Ouaftouh, Zellou, and Idri 2015), (Calegari and Pasi 2013), (Ghosh 2016) and (Golemati et al. 2007), user profile offers an organized construct containing information about a user's cognitive skills, intellectual abilities, intentions, learning styles, preferences, behavior, context, and personal data. In other words, a user profile outlines the role that the user has assumed as well as the rules, limitations, and permissions that the context establishes for each user.

User Profiling is the process of Extracting, Incorporating and Detecting the keyword-based data to create a structured profile and then visualizing the knowledge derived from these results (Elallioui and El Beqqali 2012), (Godoy and Amandi 2005),(Kanoje, Mukhopadhyay, and Girase 2016), (Kanoje, Girase, and Mukhopadhyay 2015) and (Alaoui, Idrissi, and Ajhoun 2015).

A user model can take on as many different forms as there are reasons for which they are created. User models may aim to describe, the following (Webb, Pazzani, and Billsus 2001):

- The cognitive processes that underlie the user's activities;
- The deference's between the user's skills and expert skills;
- The user's behavioral patterns or preferences; or
- The user's attributes.

There are three primary methods for defining a user profile as highlighted in (Luna et al. 2015) and (Eyharabide and Amandi 2012):

- Collect relevant information about the user preferences;
- Choose a formal language to describe this knowledge;
- Determine a strategy to update the user profile.

User profiling involves two crucial aspects: effectively getting to know users so that it can respond to their individual needs, and using that information to recommend items that could be of interest to them (Kanoje, Girase, and Mukhopadhyay 2015),(Skillen et al. 2014). Gathering data regarding users and their interests is the primary purpose of user profiling.

3.2. User profiling modeling techniques:

A user model is created and updated through the process of "user modeling," that involves deriving user characteristics from user data, which can either be information that the user actively provides or data that is derived via unintentional observations and events.

There are several ways to build a user model: ***The overlay model*** which highlights the user's knowledge level. ***The stereotype model*** which divides users into groups based on their common attributes (Rich 1979), (Chrysafiadi and Virvou 2013), (Zapata-Rivera and Arslan 2021). ***Vector representation*** to represent the profile taking into consideration the diverse interests and their evolution through time (Bouneffouf 2013). ***Machine learning techniques*** for automated induction and observation of user activities and behavior (Webb, Pazzani, and Billsus 2001),(Eke et al. 2019), (Bedi et al. 2022). ***Fuzzy logic and Bayesian network*** for addressing user diagnosis uncertainty and ***Ontologies*** for reused user models (Bouneffouf 2013).

3.2.1. Overlay model:

Overlay model is the most important and most popular explicit user modeling approach in adaptive hypermedia. It represents the user's knowledge level. An overlay model is a structural knowledge model. It serves as a subset of the domain model to describe a person's level of expertise. The knowledge status of a certain user is represented as a subset of expert-level knowledge. By using this technique, it is possible to determine how well a user can apply a concept or what their likelihood of knowing the answer to a certain problem is. Domain model and overlay knowledge model are the two required minimum components of an overlay model (Nguyen and Do 2009), (Ahmadaliev et al. 2019).

3.2.2. Stereotype model:

Stereotype model is one of the earliest techniques for user modeling. The model was first presented by Rich (1979) in the system called GRUNDY (Rich 1979) . Stereotypes were frequently used in early adaptive systems (between 1989 and 1994). A stereotype is a collection of regularly occurring attributes of users. The basic goal of stereotyping is to divide all potential users of an adaptive system into different categories based on the characteristics they frequently have in common. Such groups are called stereotypes (Chrysafiadi and Virvou 2013). If any of a new user's characteristics (such as preferences, prior knowledge, or types of errors) are similar to those in a stereotype, the new user will be placed in that stereotype. Several default assumptions are inferred from the little amount of initial information. The underlying assumptions are changed as more data regarding specific assumptions becomes available. To begin the learner model, one can employ these stereotypes. They can therefore act as a method for resolving the cold-start issue (i.e., not having enough information about the learner to implement adaptive features) (Zapata-Rivera and Arslan 2021).

3.2.3. Vector representation:

Vector spaces are frequently utilized for information organizing and retrieval. They are simpler to maintain as there are no pre-assigned groupings or stereotypes to take into account (Mangina and Kilbride 2008). The vector representation is based on the traditional vector space model of Salton et al and it was the first model of the user's profile to be used (Salton, Wong, and Yang 1975). The user's profile is represented as an m-dimensional vector, where each dimension is a separate term and m is the total number of terms in the user's profile. Typically, a TF/IDF diagram serves as the foundation for term weighting. The weight assigned to each term indicates how significant it is to the user's profile (Bouneffouf 2013).

3.2.4. Fuzzy logic modeling techniques or Bayesian networks:

Fuzzy Logic (FL) establishes a framework for modeling and reasoning in uncertain situations by capturing the inherent ambiguity of real information (Popp and Lödel 1995), (Frias-Martinez et al. 2005) . A key concept in FL theory is the notion of the fuzzy set. A fuzzy set expresses the degree to which an element belongs to a set. The degree of truth in fuzzy logic can take continuous values between [0,1 as opposed to classical binary or multi-valued logic, which draws its values from a discrete finite set. This characteristic enables the capture of the uncertainty present in real data (Chrysafiadi and Virvou 2015), (Hoblitzell, Babbar-Sebens, and Mukhopadhyay 2016). In other words, fuzzy logic allows for the analysis of verbally ambiguous input data. It permits inference and natural description of knowledge in the form of ambiguous notions, operators, and rules. (Kavcic 2004). Bayesian networks have a strong theoretical foundation, they are reliable, and they are helpful for all kinds of issues. They are particularly effective inference tools (diagnostic and predictive) but they require a

comprehensive model (variables and relations, all knowledge must be coded in the model), together with the estimations of all parameters (a priori and posteriori probabilities) (Vuong et al. 2020) ,(Rim, Amin, and Adel 2013), (Horvitz et al. 2013), (Zukerman and Albrecht 2001), (Marcot and Penman 2019).
User profiling employs various techniques, with ontology being among the most used. In the next section, we will highlight ontologies as a technique of user profile modeling.

4. Ontology-based user profile:

4.1. Ontology definition:

Ontologies provide a formal description of definitions of conceptual classes and their relations that goes beyond lists, thesauri, and taxonomies. By formal, we mean that definitions are founded on a logical framework (Harrow et al. 2019). Ontologies are used as a method for encoding the semantics of an area of human knowledge in a machine-readable form. They are essential for identifying meaningful relationships that enable users to browse or search relationships and find patterns through analysis.
Ontology is a subfield of philosophy that examines the nature and structure of "reality". It describes how things are connected naturally and how their components are connected secretly (Gao et al. 2017). In his book Metaphysics, Aristotle discussed the idea of ontology. He described it as the study of "being qua being," or the characteristics that belong to objects because of their very essence (Staab and Studer 2009). Later, it's borrowed to be used in Computer Science fields as being an effective and powerful concept semantic model.
Ontology was first described by Gruber as a «explicit specification of a conceptualization» in 1993, (Staab and Studer 2009). In this context and according to (Al-Yahya, George, and Alfaries 2015):

- A conceptualization is an abstract representation of a certain part of the world that takes the form of a definition of key terms' characteristics.
- An explicit specification means the model must provide meaning for the vocabulary while also making it machine and human processable.

This notion was expanded upon by Borst in 1997, who described an ontology as a "formal specification of a shared conception" (Staab and Studer 2009). This means that an ontology needs a formal language to explain its specification and that its definitions and meaning should be shared among groups.
An ontology offers a semantic foundation for describing digital content in a way that is machine-understandable. It is widely used in information systems, by adding meta-data to documents, enhancing information retrieval and reasoning, and facilitating data sharing between applications (Maedche and Staab 2001), (L. Zhou 2007).
Ontologies are formal models that, in addition to providing an accurate description, the logic of the meaning of the term, the data structures, and other aspects to model the real world, depict how a human perceives an area of interest (Luna et al. 2015). Moreover, according to (Corcho, Poveda-Villalón, and Gómez-Pérez 2015), ontology is also a representation of the types of entities, their relationships, and their constraints. . It consists of a collection of hierarchically arranged Classes, Relations, Instances, Functions, and Axioms.(Calero, Ruiz, and Piattini 2006).

Ontologies, which enable the representation of knowledge in a structured and expressive manner, have lately been employed as a highly expressive schema to represent user profiles (Luna et al. 2015). Ontological profiles gather user interests and offer recommendations based on these in order to maintain a tailored searches.
Modeling the user profile based on ontology has already been advocated in many applications including web search and personal information management (Golemati et al. 2007). The user profile can be created explicitly by asking the user for the relevant information about themselves, or implicitly by observing the user's behavior (Duong et al. 2009).

4.2.Ontology types:

Ontology categorization is related to the subject of the conceptualization (Gómez-Pérez, Fernández-López, and Corcho 2006). This classification organizes ontologies into four types as shown in Figure 3:

Figure 3 Classification of Ontology

This four types are described in (Al-Yahya, George, and Alfaries 2015), (Slimani 2015) and (Tarus, Niu, and Mustafa 2018a) , as follow:

- **Upper/Top level ontologies:** They represent fundamental concepts (such what time and space are) that are not specific to a given domain. They define basic concepts such as objects, relations, events, processes, and so on.
- **Domain ontologies**: are reusable within a certain domain (such as medicine, biology, tourism, movies, sports, history, or law); they represent knowledge and offer terminology for concepts and activities inside a particular domain. They are not dependent on any one particular activity or application, which is their fundamental feature.
- **Task ontologies:** describes the terminology associated with a general task or activity, such as buying, selling, or solving problems. Any domain can use the task.
- **Domain-Task ontologies** represent the vocabulary for a certain task. They cannot be used in different domains.

- **Application ontologies**: are application-dependent and define terminology pertinent to a particular application. These are specialized domain ontologies that depict a specific model of a domain from the perspective of a single user or developer.

4.3.Ontology components and languages:

4.3.1. Ontology Components:

Ontologies mostly consist of the following significant components as highlighted in (Khadir, Aliane, and Guessoum 2021),(Grimm et al. 2011) and summarized in Table 3:

Table 3 Ontology Components

Ontology Components	Description
Classes	Types of objects that determine a group of concepts. They describe the elements of the ontology.
Instances (individuals)	Objects that define the ontology enrichment. They are instantiations of classes. For example, Hajar_Zankadi is an instance of the class Learner.
Attributes	Properties attached to concepts or classes, used to assign datatypes and values to objects (e.g., Strings, Numbers).
Relations (relationships)	Connections that specify how closely two objects are related. Taxonomic and non-taxonomic relations are the two categories into which they are typically classified.
Taxonomic relations	"is-a" or "sub-class-Of" relationships representing the hierarchy of entities (e.g., Learner has-part: Learning-Style).
Non-taxonomic relations	Links expressing other types of connections, contributing to semantic enrichment without changing structural elements.
Axioms	Formal definitions of ontological knowledge expressed using First-Order Logic or Descriptive Logic. Categorized into Instantiation, Assertion, and Subsumption axioms, specifying instance-class relationships and property values.

4.3.2. Ontology languages:

Ontology languages of many kinds have been created and used, particularly in the context of the Semantic Web. Ontology representation languages like Web Ontology Language (OWL), Resource Description Framework (RDF), DARPA Agent Markup Language (DAML), and Ontology Inference Layer (OIL) are used to encode information about particular topics as presented in Figure 4:

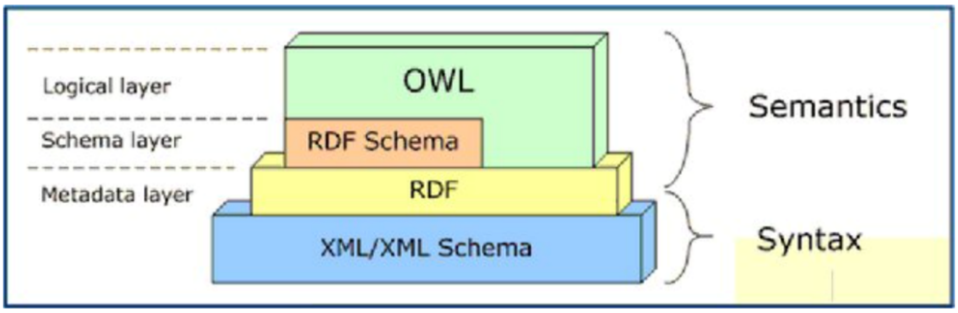

Figure 4Layered architecture of ontology languages (Panchanathan, Lawan, and Rakib 2015)

The formal languages most frequently employed for creating ontologies are OWL and RDF:

- **RDF** (F. Zhang et al. 2019):Stands for Resource Description Framework, was created by the W3C to define Web resources and permits the definition of the semantics of data based on XML in a homogeneous, interoperable manner. It offers a graph-based data model or framework for organizing data as claims about resources.
- **OWL** (McGuinness and Van Harmelen 2004): Stands for Web Ontology Language, was developed in 2001 by a W3C working committee. OWL fosters collaboration and sharing among applications, offers a wide range of operators for creating concept descriptions, and was created to be compatible with current web standards.

4.4.Ontology techniques:

4.4.1. Ontology mapping:

Ontology mapping is an essential technique for providing interoperability and heterogeneity between systems and services using ontologies (Kaladevi Ramar and Geetha Gurunathan, n.d.). Ontology mapping (or matching) is essential to afford semantic access across aggregated data used in knowledge-based products and services consumed by life science companies, academic institutions, and universities (Harrow et al. 2019).
Given two ontologies A and B, mapping one ontology with another means that for each concept (node) in ontology A, we try to find a corresponding concept (node), which has the same or similar semantics, in ontology B and vice versa."(Ehrig and Sure 2004)
Ontology mapping is classified into the following three categories, as described in (Choi, Song, and Han 2006):

- **Mapping between an integrated global ontology and local ontologies:** describes how a global integrated ontology and local ontologies are related to promote ontology integration.
- **Mapping between local ontologies:** promotes interoperability for highly dynamic and distributed environments as a mediator between distributed data in such environments.

- **Mapping on ontology merging and alignment:** is utilized as part of the ontology reuse process. It allows for the creation of a single cohesive merged ontology via an ontology merging procedure. Additionally, it establishes connections between local ontologies while they are still distinct during the ontology alignment procedure. Mappings do not exist between a single coherent merged ontology and local ontologies, but rather between local ontologies to be merged or aligned.

4.4.2. Ontology merging:

"Coming together" or "becoming one" is what is meant by the word "merging". In merging, two original ontologies, O1 and O2, are combined during merging to produce a single merged ontology. The original ontologies include related or overlapping (sub) domains but they are unique and not revisions of the same ontology (Chatterjee et al. 2017).

The new ontology will take the place of the original ones and incorporate the knowledge from all combined ones (Robin and Uma 2010). Merging frequently makes use of a set of alignments in order to establish strong linkages between ontologies and ultimately combine them into one.

5. Problem statement:

MOOCs have grown in popularity since 2008 in the field of online education. Universities and colleges provide flexible, top-notch online courses that can be taken by learners with a variety of needs and learning preferences. With no restrictions on class size, necessary skills, or severe registration requirements, MOOCs enable learners to engage in open, public online communities. They are often free of charge, with the exception of when a completion certificate is given (C. Jin 2021; Goopio and Cheung 2021).

During the COVID 19 pandemic, MOOCs received increased attention and are now regarded as a very efficient and adaptable learning method. Recently, over 2.8K courses were added, and 16.3K MOOCs were announced or launched by approximately 950 universities worldwide, demonstrating the high demand for MOOCs (Er-Rafyg et al., n.d.). According to (News, n.d.), over 180 Million learners have enrolled MOOCs' courses as highlighted in figure 5.

New Registered Users	2019	2020	Total
coursera	8M	31M	76M
edX	5M	10M	35M
FutureLearn	1.3M	5M	15M
class central	350k	800k	2.3M

Figure 5 Learners' enrollment in MOOCS during COVID 19 pandemic (News s. d.)

Chapter 1 Context and background of social e-learning and learner profiling

According to data made public by Class Central, MOOC enrollment increased by one-third, reaching 180 million learners globally (excluding China). The main MOOC providers, Coursera, edX, Udacity, and FutureLearn, launched over 360 micro-credentials and 19 online degrees in 2020. Many of them offered complimentary certificates.

However, several issues faced by MOOCs have come to the fore, notably the high dropout and low retention rates that occurs in the early stages of learning (J. Chen et al. 2019) . Dropout in MOOCs occurs when learners fail to finish the course (Dass, Gary, and Cunningham 2021). According to (Badali et al. 2022), around 90% of enrolled learners drop out before the course is over, and the retention rate for MOOCs is between 3 and 15%.

Several factors, including lack of continuous motivation and engagement, less interaction, learner characteristics, MOOC features, and learner experience, have an impact on the decisions made by learners at all stages of MOOC acceptance (Badali et al. 2022) (K. P. Gupta and Maurya 2022) (Panagiotakopoulos et al. 2021) (Ji, Park, and Shin 2022).

Table 4, summarizes the key factors affecting learners' decision to drop-out in MOOCs. The factors vary from learners-related factors such as learners' characteristics and motivation and MOOC-related factors such as unsuitable courses design.

Table 4 Factors that influence learner dropout in MOOCs

Reasons	Factors	References
Social interaction	- Lack of social interaction and cooperative activity between the learners and group work - learner assistance: (interaction with course content, peers, instructor) - Engagement - Forum participation - Interactive activities - Social influence - Sense of community	(El Said 2017); (Rosé et al. 2014); (D. Yang, Wen, and Rose 2014); (Zheng et al. 2015) (Feng, Tang, and Liu 2019) (Itani, Brisson, and Garlatti 2018) (Wang et al. 2019); (Q. Zhang et al. 2016)
Learner profile	- Heterogeneity of the learners' profile - Personal characteristics	(J. Chen et al. 2019) (Gütl et al. 2014); (Khalil and Ebner 2014); (Shapiro et al. 2017)
Motivation	- Lack of motivation on the part of the learners - The desire to earn a certification - Level of engagement - Social situation and characteristics	(Eriksson, Adawi, and Stöhr 2017); (Gomez-Zermeno and de La Garza 2016); (Y. Zhou, Zhao, and Zhang 2020) (Zheng et al. 2015)

Course design	- Course structure - Course content - Course duration - Course quality - System and service quality - Duration of video lectures	(Gomez-Zermeno and de La Garza 2016) (Feng, Tang, and Liu 2019) (Itani, Brisson, and Garlatti 2018) (Shapiro et al. 2017) (Shawky and Badawi 2019),
Technology	- Use of web technology - Innovative features - Task-technology fit - Mobility Internet access	(El Said 2017); (Gomez-Zermeno and de La Garza 2016);
Language	- English proficiency - Complex sentences - Sophisticated terms - Unclear accent of instructor	(El Said 2017); (Eriksson, Adawi, and Stöhr 2017); (Gomez-Zermeno and de La Garza 2016)
Time	- Time devoted to MOOCs - User active time - Ability to manage time - Time zone differences	(Eriksson, Adawi, and Stöhr 2017); (El Said 2017) ,(Gomez-Zermeno and de La Garza 2016) (Feng, Tang, and Liu 2019) (J. Chen et al. 2019) (Shapiro et al. 2017)

According to Table 1 and to other works, (Bezerra et al. 2017), (Eriksson, Adawi, and Stöhr 2017), (Aldowah et al. 2019), and (Goopio and Cheung 2021), the main reasons that influenced the decision to dropout are articulated into four main factors as shown in Figure 6:

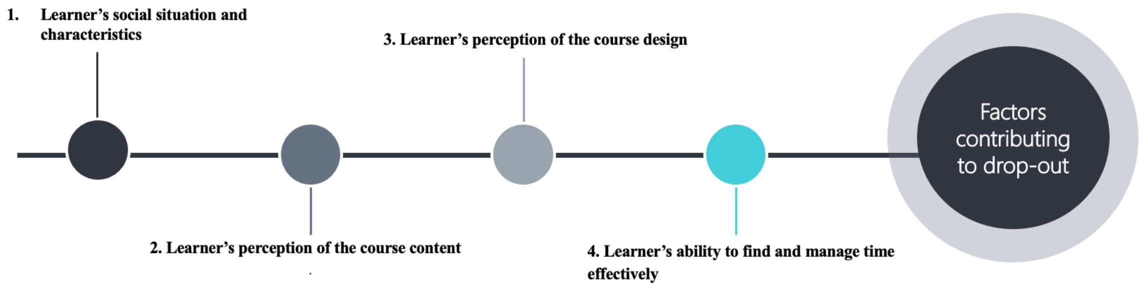

Figure 6 Main Factors contributing to drop-out in MOOCs

We are interested in the learner's characteristics and profile, social situation, and the impact of social elements on dropout decisions in MOOCs. we explore the crucial role of learner engagement, satisfaction, and the influence of social media in the e-learning context. Table 5 presents a comprehensive overview of the factors related to learners ‘characteristics and social integration impacting the dropout rate in MOOCs:

Table 5 Comprehensive overview of factors impacting dropout rate and related to learner ‘s characteristics and social integration

	Factors	**References**
Learner ‘s characteristics	▪ Learner profile (e.g., Interest, preferences, learning styles, etc.)	(Gütl et al. 2014); (Khalil and Ebner 2014); (Shapiro et al. 2017)
	▪ Learner ‘s motivation ▪ Learner ‘s satisfaction	(Xiong et al. 2015); (Khalil and Ebner 2014) (J. Han, Geng, and Wang 2021) (Alqurashi 2016)
Social Integration	▪ Struggles of learner to fit into the social community ▪ Struggles of learner to adhere to the academic norms	(Lee and Choi 2011)
	▪ Social presence ▪ Engagement	(Rosé et al. 2014); (D. Yang, Wen, and Rose 2014); (Zheng et al. 2015) ;(Wang et al. 2019); (Q. Zhang et al. 2016)
	▪ Social media's interactivity	(Lin and Kishore 2021)
	▪ Meeting social requirements	(Kross et al. 2021)

Chapter 1 Context and background of social e-learning and learner profiling

Learner engagement and satisfaction drive MOOC success(J. Han, Geng, and Wang 2021) (Alqurashi 2016), supported by social learning's efficacy in reducing isolation. Social e-learning refers to the development of either online learning platforms with social media features or social media sites used for educational purposes (Troussas, Krouska, and Sgouropoulou 2021) . The use of social media in e-learning enables the learning to occur without being constrained by physical locations and in a variety of creative ways, including social interaction, online collaborations, customized and user-friendly multimedia interface that offers a high level of participation, as well as cooperation and contact between users (Krouska, Troussas, and Virvou 2019) (Hajli et al. 2013). Additionally, social media can greatly enhance learner engagement in the learning process by promoting interaction between learners and teachers, feedback on course materials in the form of likes and comments, motivation for learners in the form of notifications, the formation of cohesive learner groups for collaborative tasks, among other things (Alalwan et al. 2019).). Studies have shown that social media provide advantageous opportunities for teaching and learning in the light of the promoted constructivist approaches. Which means that learning happens when learners actively contribute to the process of knowledge building rather than only absorbing information (Gray 1997).Thus, social media was highlighted as having the ability to boost networking and engagement between teachers and learners (Greenhow and Askari 2017) and are considered a powerful driver of change for learning practice, in terms of openness, interactivity and sociability (Manca and Ranieri 2016). After reviewing the factors behind drop-out in MOOCs using literature and highlighting the potential of using social media information and its features to promote learner engagement and satisfaction, we aim to address the problem of drop-out by cooperating the information of both environments (MOOCs and social media).

The aim of our thesis is to exploit the learner ‘s information and personal involvement in social media for the purpose of enriching and personalizing his experience in MOOCs. As we know, different learners in MOOC have big differences in learning behaviors, learning habits, and learning time, etc. which leads to different learner profiles and learning needs. The objective of this thesis is leveraging social media for personalized MOOCs experience through Learner Profiling and Interest Detection using ontologies and NLP.

6. Conclusion:

This chapter emphasized the numerous concepts and methods we employed for our first contribution that aims to build a social profile for learners within MOOCs using ontology engineering techniques in order to enrich their profile and foster their engagement for a more personalized learning experience.

We began by highlighting the evolution of the term "e-learning," that has transformed education by leveraging technology and the internet to provide learners with flexible and accessible learning opportunities. It encompasses various definitions, including the use of technology for learning and the integration of electronic media in education. Supported by software programs like Learning Management Systems (LMS), Course Management Systems (CMS), Learning Content Management Systems (LCMS). We focused as well on Massive Open Online Courses

(MOOCs) that offer a wide range of resources and platforms for learners to engage with educational content with no limitations regarding time and location.
The social learning theory was then highlighted, along with its benefits for e-learning when combined with social media. It points out the importance of observation and modeling in the learning process. In the context of e-learning, social media has emerged as a powerful tool for facilitating social learning. It encourages active engagement, interactions, and knowledge sharing among learners, leading to increased self-efficacy and improved learning outcomes. Social networking platforms enable the generation and sharing of diverse content, allowing learners to emulate actions and learn from other users' problem-solving strategies.
Finally, we reviewed various methods for user profiling that plays a vital role in understanding learners' characteristics, preferences, and behavior. By creating structured profiles that encompass cognitive skills, learning styles, and personal data, user profiling techniques enable personalized learning experiences. Ontology-based user profile modeling, using ontologies to represent user interests, further enhances personalization and enables the provision of tailored recommendations. Ontologies provide a formal and semantic foundation for representing knowledge, facilitating the understanding and analysis of user data.

In the next chapter, we point out the main works that have been done in the area of social e-learning, learner profile using ontologies and interest detection using NLP techniques.

Chapter 2: Related Work

Chapter 2: Related Work

Introduction:

In this chapter, we explore the impact and challenges related to the use of social media on learning. We examine their collaborative impact on learners. We investigate the learner characteristics through learner profiling exploring the advantages of using ontology approaches to model the learner profile through structured framework for a better understanding of learners needs. We also highlight the use of topic modeling techniques in the context of both MOOCs and social media to uncover the hidden aspects in the textual content.

1. **The impact of the use of social media on E-learning:**

With the remarkable rise of e-learning during the COVID-19 pandemic, education has undergone a tremendous alteration. This change was necessary because colleges and universities were closing. The use of social media as a tool for learning has received much attention, particularly among learners who can connect and work together in a variety of group interactions (Kolokytha et al. 2015). Its opportunities for communication, resource sharing, and promoting active learning and critical thinking had greatly helped in integrating social media features into the educational process.

Due to the site's simple and practical interactivity features, such as liking, commenting, polling, and sharing content, learners are more likely to use social media in the context of higher education to increase knowledge sharing, behaviors, and improve their learning performance as opposed to other conventional learning approaches. Any e-learning system has a built-in social network, with teachers, learners, and learning resources as its core actors. With the utilization of its characteristics, social media has gradually transformed from a digital hub for social communication into a commercial, educational, and social regulatory institution. The availability and use of social media is growing due to reasons such as enhanced software tools, powerful computers, and mobile devices, as well as increased access to broadband services (Alghizzawi et al. 2019).

Social media has a significant impact on the learner community nowadays and is gradually integrating into everyone's daily lives in contemporary culture. According to statistics, 55% of people worldwide actively use social media ("Digital 2021 April Statshot Report," n.d.). It is distinguished by its social approach, which principally relies on interpersonal interactions as the basis for knowledge co-construction and which helps to change society norms, values, and culture. (Chukwuere and Chukwuere 2017) (Hardy and Castonguay 2018).

Given the expanding popularity of both social media and e-learning, it seems logical to combine these two popular technologies in order to enhance online teaching and learning (Zankadi et al. 2018). The SWOT analysis highlights the motivating factors for social media adoption in e-learning.

Defining the SWOT means identifying the Strengths, Weaknesses, Opportunities, and Threats that affect a company and its performance, which is a process carried out by the management team(Leigh 2009). Strengths in the SWOT analysis refer to internal resources and positive aspects of business establishment that are important for businesses to successfully accomplish their goals and provide effective customer service (Gurl 2017). Internal obstacles or limitations are referred to as weaknesses and might affect an organization's success. As a result, internal factors are what make up the company's strengths and weaknesses.(Teoli, Sanvictores, and An 2022). The SWOT analysis identifies opportunities for business establishments with links to external organizations based on factors or features that favor or facilitate them (Leigh 2009). whereas threats are external factors that can obstruct or delay the company's achievement of its objectives.

Table 6 SWOT analysis of using social media in e-learning

Strengths	Weaknesses
▪ Personalized learner experience. ▪ Learners develop communication and interpersonal skills. ▪ Enhances social constructivist technique to learning. ▪ Promoting a sense of belonging among learners. ▪ Developing communities of practice. ▪ Facilitate the sharing of information. ▪ Promote social presence which is crucial for education and learning. ▪ Through social media, learners can construct virtual communities of practice that provide a global platform for collaboration and social interaction (Brady, Holcomb, and Smith 2010). ▪ Through shared readings, links, and videos, learners build a shared understanding and collaborate in discussion. ▪ Aggregation of learners with common interests. ▪ It has an advantage for learning foreign languages and help learners master both their written and oral languages proficiency.	▪ Concerns about learner privacy and safety influence the decision of many education administrators to ban the use of social media in the learning process. ▪ Copyright issues. ▪ Evaluate originality. ▪ Sensation of informational restraint. ▪ Sometimes, instructors unintentionally omit to provide the materials needed to aid the student's learning process. (Al-Shawabkeh and Lim 2014). ▪ Social Media is time consuming. ▪ False identity creation. ▪ Fraudulent use of identification. ▪ Lack of trust in social media content prevents learners from feeling comfortable using it.

Opportunities	**Threats**
▪ Self-development of learners can be done through social media platforms. ▪ Social media encourages continuous learning as it saves times and money. ▪ The social media integration technology enables students to learn new information by reading online materials and chatting with instructors whenever and wherever they like. ▪ Social media can enhance writing abilities and reading habits due to the availability of free books online. ▪ Learners can motivate other leaners to speak and write foreign language more confidently. ▪ Social media platforms, which are engaging and appealing, can help learners develop their vocabulary. ▪ Due to the frequent updating of material on social media, there will be a desire to learn more. ▪ With social media, learners develop an automated cognitive system that produces behavioral and motivational impulses to engage in rewarding actions. This system allows for the mental registration of links between inputs, behaviors, and rewards. (Meshi, Tamir, and Heekeren 2015)	▪ The total reliance on social media and the easy availability of information makes learners unproductive. ▪ Social media could be a distraction and waste of time. ▪ A lack of knowledge about social media content can lead to choosing the wrong resource. ▪ Social media can be an addiction. ▪ The use of social media might increase stress. ▪ Concerns over morality and privacy. ▪ Ownership issues. ▪ Workload issues. ▪ Fraud and scams.

Table 6 highlights various strengths, weaknesses, opportunities, and threats associated with the integration of social media in the learning process. Strengths include the promotion of personalized learner experiences, development of communication skills, and facilitation of social constructivist techniques. Opportunities encompass self-development, continuous learning, and enhancement of writing abilities through engaging platforms. On the downside, concerns about learner privacy, copyright issues, and the potential for distraction and addiction pose challenges. The comprehensive evaluation suggests a nuanced perspective on the multifaceted impact of social media in education, necessitating careful consideration of both advantages and drawbacks in its implementation.

Currently, social media is widely used in MOOCs due to its many advantages, including its flexibility and convenience (Giannakos, Mikalef, and Pappas 2022). Social media has become a tool that can complement learning.

Table 7 presents the most recent studies that point out the use of social media in MOOCs, focusing on the main finding of each work:

Table 7 The use of social media in MOOCs

Title	Main Finding	Reference
Public perceptions towards MOOCs on social media: an alternative perspective to understand personal learning experiences of MOOCs	▪ Weibo is used as a public service medium to increase the accessibility of MOOC portals and as a space for learners to share their personal learning experiences, reflecting aspects of autonomous, self-regulated, interactive, and cooperative learning. ▪ MOOCs provide learners with autonomy and the capacity to express themselves informally and talk about their learning. - Weibo facilitates the formation of serious learning communities or groups through close peer connections and continual interactions to assist MOOC learning.	(Shao et al. 2023)
Pattern of social media engagements by the learners of a library and information science MOOC course: an analytical study	▪ Learners' active participation in the online discussion forum increased over time. ▪ The performance of social media involvement, particularly through the YouTube channel, became popular. ▪ The paper suggests the potential development of stable performance indicators based on online engagements in LMS platforms.	(Naskar, Hasan, and Das 2021)
An evaluation of social learning and learner outcomes in a massive open online course (MOOC): a	▪ Social media usage does not significantly affect knowledge outcomes but enhances affective outcomes. ▪ Engagement with social media tools does not impact knowledge-related learner outcomes, but	(Anderson, Gifford, and Wildman 2020)

healthcare sector case study	it provides a basis for enhanced attitudinal or 'affective' learner outcomes. ▪ The use of social media tools in the MOOC program did not significantly impact knowledge-related learner outcomes, but it did contribute to positive affective learner outcomes.	
How can the participation in massive open online courses be increased? impact of social media	The main findings include the investigation of reasons for the preference of MOOCs, the importance of social media in interaction, and the disadvantages of MOOCs.	(Bicen and SAPANCA 2020)
Gamification of MOOCs Adopting Social Presence and Sense of Community to Increase User's Engagement: An Experimental Study	▪ The main findings are that the game elements successfully triggered social presence and SoC among MOOC users, but the expected association with user engagement was not confirmed. ▪ The study should be viewed in light of several limitations, including a limited sample size and potential bias in the questionnaire data.	(Antonaci et al. 2019)
Modelling an interplay of adoption determinants with respect to social Web applications used in massive online open courses	The main findings of the paper are the potential benefits of employing social Web applications in MOOCs to boost students' motivation, the influence of students' acceptance on the successful implementation of social Web applications in MOOCs, and the examination of the psychometric characteristics of the research framework reflecting the interplay of adoption determinants with respect to two representatives of social Web applications meant for collaborative work.	(Orehovački, Etinger, and Babić 2019)
Exploring Social Media Data for MOOC Recommendation	The main findings of the paper are the focus on using social media mining to acquire personal and professional data about learners for MOOC recommendation, addressing the lack of motivation in MOOCs, and outlining the process of applying data mining on a social media dataset for MOOC recommendation.	(Assami, Daoudi, and Ajhoun 2019)

The convenience and interactive features of social media make it a preferred tool for knowledge sharing and improving learning performance. Social media's growing popularity and its impact on contemporary culture make it a logical choice for integration with e-learning to enhance teaching and learning experiences. SWOT analysis highlights the motivating factors for adopting social media in MOOC, considering its strengths, weaknesses, opportunities, and threats. Social media offers flexibility, convenience, and the ability to complement learning in MOOCs, creating engaging learning environments and facilitating knowledge sharing. It also positively influences learner engagement, academic achievement, interpersonal skills, and self-confidence. The use of social media platforms like Facebook and WhatsApp has shown promising results in terms of learner satisfaction, interaction, and creating a sense of community. Overall, social media has transformed how learners engage, communicate, and socialize in their academic careers, making it a valuable tool in the e-learning landscape.

MOOCs and social media differ in terms of user interaction. This diversity is important because social media interactions are so powerful and spontaneous that they help uncover hidden information about user profiles. Determining the user's profile elements help to get a bird's eye view into what skills, interests and preferences matter for different learners in e-learning environment. It is crucial to make learners engaged in their learning process which can be achieved by embracing personalization and socialization. This can ensure that the learning efforts are aligned with what responds with their profile requirements.

Using social data to construct the learner profile is known as learner social profiling. By doing so, it will be possible to learn more about the needs of the students and which one has to be strengthened.

The aim is to make better use of learner interactions and generated content on social media to further facilitate interaction in MOOCs through ontology-powered profiles which will be highlighted in the next section.

2. Learner profiling:

2.1. Types of learner profile:

The learner profiling system resolves issues by creating, managing, and displaying changed information about each learner. Learner profiling is essential in the e-learning field for satisfying and meeting the diverse needs of learners (Rezgui, Mhiri, et Ghédira 2014).

There are three types of learner profile as presented in Table 8 (Cufoglu 2014), (T. Kulkarni, Kabra, and Shankarmani 2019), (Eke et al. 2019):

Table 8 Types of Learner Profile

Type	Description	Techniques used	Advantages	Disadvantages
Explicit	User manually creates user profile. It is also known as static profiling that is based on the static and predictable features of the user.	Questionnaires, Rating By examining the information collected from online forms, surveys, and other forms of user-	The data obtained is often of excellent quality.	Updates to the user's profile information are labor-intensive.

		submitted data, one can forecast how users will behave. Via forms or surveys, an e-Learning user can be questioned about their interests, knowledge, and other factors.		
Implicit	The system creates a user profile based on the interactions that have occurred between the user and the material in the past. It is a dynamic profiling approach. This type dynamically records or collects implicit user information. It does this by observing or tracking user interactions with the system. This approach examines user behavior to identify their interests. The user's visits to various links, the content they liked or downloaded, and other information can be tracked in the case of an e-learning site to improve the system's recommendations.	Machine learning algorithms	Easy to update in an automated way with minimal user effort	Requires a large amount of user content interactions first before creating accurately a user profile
Hybrid User Profiles	User profiles that combine explicit and implicit information. Incorporates both the advantages of explicit and implicit profiling. It takes	Both explicit and implicit techniques	to strengthen each of the methods adopted and decrease their weaknesses	N/A

	into account both static features and user interactions with the system, which dynamically updates the user's profile.			

Learner profiling is crucial in the field of e-learning to meet the diverse needs of learners. There are three types of learner profiles: static profiles, which contain unchanging user attributes and available content; dynamic profiles, which capture user behavior and interests over time; and hybrid profiles, which combine static and dynamic elements. Static profiles rely on user-provided information but can be unreliable due to privacy concerns. Dynamic profiles leverage temporal behavior analysis to capture users' evolving interests and preferences. Researchers have explored various approaches to learning dynamic profiles, such as using optimization techniques and graph Laplacian regularization to identify wellness events in users' timelines, and employing models like streaming keyword diversification and dynamic user and word embedding to track semantic representations and distinguish long-term and short-term interests.

These efforts contribute to the development of more accurate and personalized learner profiles in e-learning environments. We will discuss ontological techniques for learner profiling and related research in the section that follows.

2.2. Learner profiling using Ontologies:

The process of creating learner profiles is crucial to personalize the learning experience. We provide learners with content based on his profile information. A learner profile can be developed by asking learners directly for information or by observing how they behave within the environment. Learner profiles are a valuable source of data since they not only include the basics information, such the name, age, and gender of the learners, but also information on their learning abilities, learning preferences, interests, and conditions (Rezgui, Mhiri, and Ghédira 2014). Identifying the learner information and structuring the information in a way that facilitates semantic retrieval are the major challenges in the creation of learner profiles.

The ontology representation of the user profile, particularly in MOOCs, has the ability to define the terms used to explain and express information in the learner's profile. It allows for the reuse of domain knowledge and the exchange of understanding among users. Additionally, based on an accepted model, it analyzes, separates domain information from operational knowledge, expressly shares, and exchanges profiles throughout the system.
Numerous works have highlighted the learner profiles creation using ontologies in various fields as presented in Table 9:

Table 9 Overview of recent works in ontology-based learner profile

Reference	Description	Techniques / Methods	Key Contributions / Findings
(Assami, Daoudi, and Ajhoun 2023)	Semantic-recommendation approach for enriching learner profiles in MOOCs based on behavior and motivation in a professional social network (LinkedIn).	Semantic Recommendation, MOOCs, LinkedIn Data	Utilization of LinkedIn data for instantiating a learning actor ontology and matching learner needs with MOOC characteristics.
(Peter Ozioma et al. 2022)	Development of OntoULP, a semantic web ontology-based e-Learning management system.	OntoULP, Semantic Web, PHP, MySQL, JavaScript	Creation of a semantic web ontology-based e-Learning management system for improved search analysis and retrieval of tailored information.
(Kordahi 2022)	Presentation of OntoULP, an ontology and application for building digital User-Learner Profiles.	OntoULP, IMS LIP, Bloom's Taxonomy, Knowledge Domains Ontology	Utilization of OntoULP for constructing tailored interactions and retrieving information based on IMS LIP, Bloom's Taxonomy, and knowledge domains ontology.
(Missaoui and Maalel 2021)	Development of a learner model and categorization of profiles in an adaptive gamified learning system.	Ontology, Fuzzy Clustering, Classification Models	Utilization of Fuzzy Clustering and Classification Models for learner profiling in an adaptive gamified
(Ouatiq et al. 2021)	Creation of an ontology-based learner profile incorporating pedagogical and psychological traits along with COVID-19 health risks.	Ontology, Knowledge Engineering, IMS-LIP Standard	Development of an ontology-based learner profile combining pedagogical and psychological traits with COVID-19 health risks.
(Bennani, Maalel, and Ghezala 2020)	Proposal of an adaptive gamification domain knowledge ontology called "AGE-Learn."	AGE-Learn Ontology, Gamification, Personality Traits	Introduction of AGE-Learn ontology for incorporating learner's experience, personality traits, and motivations into gamified learning systems.

(EL AISSAOUI and OUGHDIR 2020)	Ontology-based recommender system for e-learning using Felder-Silverman (FSLSM) learning style model.	Ontology, Recommender System, Felder-Silverman (FSLSM)	Incorporation of FSLSM for modeling learner profile and learning content in a recommender system, with two ontologies used for representation.
(Kourtiche, mohamed Benslimane, and Hacene 2020)	Development of OUPIP, an ontological user model for impairment persons, enhancing personalization based on disabilities and dynamic environments.	OUPIP, Ontological User Model, Disabilities, Dynamic Environments	Creation of OUPIP as an ontological user model for personalizing software and hardware based on disabilities and dynamic environments.
(Joy and Raj 2019)	Ontology model for adaptive learner profiles and common Learning Object (LO) properties in content recommendation	Ontology, JENA API, IEEE LOM Standard	Development of an ontology model using JENA API and IEEE LOM Standard for tagging learning objects and building adaptive learner profiles.
(Bourekkache et al. 2019)	Ontological method for describing learner profiles and learning styles in personalized E-learning.	Ontology, Multi-Agent System, Learning Styles	Utilization of a multi-agent system for modular, autonomous, and interactive components in ontological modeling of learner profiles and learning styles.
(Nilashi, Ibrahim, and Bagherifard 2018)	Hybrid recommender system using Collaborative Filtering, dimensionality reduction, and ontology approaches.	Collaborative Filtering, Dimensionality Reduction, Ontology	Addressing sparsity and scalability issues in recommender systems through a hybrid approach with ontology.
(Yago et al. 2018)	Description of ON-SMMILE, an ontology network-based learner model for Multiple Learning Environments.	ON-SMMILE, Ontology Network, Semantic Web	Integration of learners' knowledge state with rubrics, objectives, and units of learning information in a semantic web-based learner model.
(Elias, Lohmann, and Auer 2017)	Usage of an ontology to address accessibility in	Ontology, IMS Global Learning	Reuse and extension of the ACCESSIBLE ontology to represent

	OpenCourseWare (OCW) systems.	Consortium Specifications, ACCESSIBLE Ontology	learner needs and preferences for improved accessibility in OCW systems.
(Sateli et al. 2017)	RDF Schema (RDFS) based model for automatically creating semantic user profiles for academic authors	RDF Schema (RDFS), Natural Language Processing (NLP)	Use of RDF Schema (RDFS) and NLP techniques for automatically creating semantic user profiles for academic authors.

The table provides an overview of various research works related to ontologies and learner profiles in the context of e-learning and recommendation systems. It covers a wide range of topics, including the development of ontologies for personalized learning experiences, the use of semantic technologies for user profiling, and the application of ontologies in gamified learning systems. The table also discusses recommender systems based on ontologies, addressing accessibility in MOOCs, and proposing methods for creating social networks based on users' knowledge interests.

Overall, learner profiling using ontologies offers a promising approach for capturing and representing learner information in MOOCs, enabling personalized and adaptive learning experiences. Ontologies facilitate the exchange and reuse of knowledge, supporting the development of intelligent recommender systems, adaptive learning environments, and collaborative learning platforms.

The idea behind building the learner profile is to customize the learning experience for each learner according to his unique profile by capturing his interests and preferences. As generating the user's interest can be challenging, topic modeling techniques are helpful in uncovering relevant thematic information related to the user. The next section introduces some works in the area of MOOCs and social media using topic modeling techniques.

3. Topic modelling:

Topic modeling techniques analyze the textual content and identify patterns, recurring themes, and underlying topics that are present in the data. By extracting and organizing the main themes from the text, topic modeling enables a better understanding of the content and allows for personalized recommendations and content filtering. In the context of MOOCs, topic modeling was used to detect learning topics interest from course reviews and discussion forums as well.

3.1. Topic modeling in MOOCs:

In the context of MOOCs, understanding what captures learners' attention has gained precision through advanced topic modeling techniques. These methods, focusing on discussion forum posts and course review, reveal the evolving subjects that engage learners.

According to (Sanya Liu et al. 2017), authors developed an author topic model, in order to extract the topic of interest for customized course recommendation. In another work (Peng et al. 2016), authors suggested a Like-Latent Dirichlet Allocation (Like-LDA) model in order to create a learner subject interest profile by including the behavioral characteristic "like". Furthermore, a Behavior-Sentiment Topic Mixture (BSTM) topic model, which combines emotion and behavior to reveal learners' focused subjects as well as learners' attitudes and behavioral patterns toward these topics, was mentioned in the work given in (Sannyuya Liu et al. 2019). Moreover, authors in (Lubis, Rosmansyah, and Supangkat 2019), created a topic model using LDA with previous knowledge relevant to the course based on constructive subjective reviews and sentiment analysis. The goal is to understand the elements that affect the teaching and learning process.

Time Information-Emotion Behavior Model (TI-EBTM), an improved dynamic topic model that automatically tracks the temporal subject changes in discussion forums, was used by authors in (Peng et al. 2020) to assess the discussion postings of learners. Similarly, Latent semantic analysis (LSA) was also utilized in the study mentioned in (B. Yang et al. 2022) to categorize and identify students who had unique longitudinal profiles of topic-relevant forum posting in MOOCs. Further, Authors in (Onan and Toçoğlu 2021), developed a document-clustering approach on discussion forum posts from MOOCs based on weighted word embeddings and clustering to identify discussion post question subjects.

3.2.Topic modeling in social media:

Social media is a valuable resource for knowledge discovery and behavior analysis. One of the most widely used social networks is Twitter, whose assessment and analysis can be highly useful for examining user behavior. Social media offers reliable sources for analyzing user activity and capturing user preferences. Using topic modeling techniques (such as LDA) can play a significant role in the discovery of hidden structures related to user behavior in social media because the user generated data (such as users' activities, user interests) in social media is a challenge.

In (He et al. 2020), authors used a probabilistic topic model to automatically identify interest tags for non-famous people on social networks like Twitter based on their social connections to famous users, without requiring text content information. Further, a Forum-LDA model is created in (C. Chen and Ren 2017) to simulate the joint generative process of the root post, relevant and irrelevant response posts. The model discovers coherent subjects and significant interests as well as unethical users who frequently add irrelevant content. Moreover, authors in (K. Kim et al. 2021), suggested a framework based on the LDA model, using both text and visual features to detect user interests and recommend POIs and potential friends to target users. In addition, a method for identifying the interests of microblog users based on multi-granular text feature representation was provided in .(Yu and Li 2021). The authors employed a text vector for microblogging, an LDA model to extract the content's topic aspects, and an LSTM technique to learn the semantic features of the phrases.

Topic modeling not only helped identify main themes in the text but also enabled a detailed exploration of sentiment, improving our capacity to uncover and comprehend the emotional aspects present in the content as highlighted in the following works:

The work pointed out in (Heidari and Jones 2020), presented a new model using Bidirectional Encoder Representations from Transformers (Google Bert) for tweet sentiment classification to find topic-independent characteristics features for social media bot detection. In (Pathak,

Pandey, and Rautaray 2021), a deep learning-based topic-level sentiment analysis model was put forth to provide scalable and dynamic topic modeling over streaming short text data in social media and to carry out sentiment analysis at the topical level. The authors in (Bhattacharya et al. 2014) concentrated on topical tweeter recommendations and provided a cutting-edge method for identifying a user's interests on Twitter. They utilized an L-LDA model (Labeled Latent Dirichlet Allocation) to identify latent topics between two tweet sets. The authors discovered that their approach to determining user interest may be better to content-based approaches.
In addition, a recommendation system based on LDA named TWILITE was proposed (Y. Kim and Shim 2014) for gathering the behaviors of Twitter users. More specifically, TWILTW has incorporated ranking algorithms to suggest users' top-K Twitter followers based on the topic distributions of users' tweets.
By analyzing the entities that users mention in their Tweets, the authors of (Michelson and Macskassy 2010) create a topic-profile that describes the topics of interest of users. The method makes use of a knowledge base to understand and classify the entities in the Tweets. In (Xu et al. 2011), the authors suggest a modified author-topic model called the twitter-user model to extract their topics of interest. The model employs a latent variable for each individual tweet to determine whether it is associated with its author's interests. In (Bhattacharya et al. 2014), authors provide an approach for theme discovery of a user on Twitter. A Labeled Latent Dirichlet Allocation (L-LDA) model is used to identify latent topics between two tweet sets. The authors of (Giri et al. 2014) employ an LDA model to comprehend the interests of the mobile users based on their browsing history. They draw attention to the fact that a user's time spent on a particular website is closely related to their interest in its contents. Based on this, they suggest an expanded model that effectively incorporates this durational information by oversampling the URLs. In (Besel, Schlötterer, and Granitzer 2016), authors described a method for determining a user's interests from Twitter followers (the accounts the user follows) rather than tweets. The English Wikipedia is used as the knowledge base for extracting named entities, and a spreading activation algorithm is used to aggregate the various interests into a more abstract interest profile. In (Piao and Breslin 2017), authors suggest a user modeling approach that uses followees' (the accounts they follow) biographies to determine user interest profiles for inactive users. By simultaneously assuming user groups and interests based on social media interactions, the authors of (Sadri, Hasan, and Ukkusuri 2019), provide a method to characterize social interaction networks. They present three main models for drawing conclusions about patterns: (1) The interest pattern model (IPM), which captures topics of interaction at the population level; (2) The user interest pattern model (UIPM), which captures topics of interaction specific to the user; and (3) The community interest pattern model (CIPM), which captures both community structures and user interests.
Topic modeling techniques have been applied in both MOOCs and social media contexts to uncover relevant thematic information related to users. In MOOCs, topic modeling is used to detect learning topics of interest from course reviews and discussion forum posts. Authors have developed various models such as author topic models, Like-LDA models, and Behavior-Sentiment Topic Mixture (BSTM) models to extract learners' subject interests, behavioral patterns, and attitudes toward topics. These models help in customized course recommendations, learner subject interest profiling, and understanding factors affecting the teaching and learning process.

In social media, topic modeling techniques are valuable for analyzing user behavior and identifying user interests. Topic modeling has been applied to social networks like Twitter to automatically identify interest tags for non-famous users based on their social connections, simulate generative processes of discussion posts, detect user interests using text and visual features, and identify microblog user interests based on multi-granular text feature representation. Sentiment analysis in social media has also been performed using topic modeling, where models have been developed for sentiment classification, sentiment analysis at the topical level, and tweeter recommendations based on latent topics.

Overall, topic modeling in both MOOCs and social media allows for the discovery of hidden structures, user interest profiling, behavior analysis, and sentiment analysis. These techniques help uncover relevant thematic information related to users, enabling personalized recommendations, understanding user preferences, and enhancing user engagement and satisfaction.

Conclusion:

In conclusion, the integration of social media platforms in MOOCs has significantly impacted the field of e-learning, particularly during the COVID-19 pandemic. Social media platforms offer numerous opportunities for communication, resource sharing, and active learning, thereby enhancing the educational process. The convenience, interactivity, and popularity of social media make it a preferred tool for knowledge sharing and improving learning performance in MOOCs.

The learner profile plays a crucial role in the field of e-learning, as it aims to meet the diverse needs of learners. Learner profiling involves capturing and managing information about each learner, including their basic attributes, learning styles, preferences, and interests. By creating and maintaining learner profiles, e-learning systems can provide personalized and adaptive learning experiences. The use of ontologies for learner profiling in e-learning systems enhances the representation and exchange of knowledge, supporting the development of intelligent recommender systems, adaptive learning environments, and collaborative learning platforms.

Topic modeling techniques play a vital role in uncovering relevant thematic information related to learners. In both MOOCs and social media contexts, topic modeling has been applied to extract learning topics of interest, identify behavioral patterns, and understand user preferences.

In MOOCs, topic modeling is used to analyze course reviews and discussion forum posts, enabling customized course recommendations, learner subject interest profiling, and a deeper understanding of factors influencing the teaching and learning process. In social media, topic modeling assists in analyzing user behavior, identifying user interests, and sentiment analysis. By utilizing topic modeling techniques, personalized recommendations, content filtering, and a better understanding of learners' needs can be achieved.

The impact of social media use in MOOCs, combined with learner profiling and topic modeling, has transformed the way learners engage, communicate, and socialize in their learning journeys. Learners' satisfaction, engagement, and academic achievements have been positively influenced by the integration of social media platforms, allowing for more interactive and collaborative learning experiences. Learner profiling, facilitated by ontologies, enables the customization of learning experiences based on individual learners' unique profiles, ensuring that learning efforts align with their specific requirements. Topic modeling techniques further enhance the personalization of learning by uncovering relevant thematic information and tailoring content recommendations to learners' interests.

Nonetheless, a lot of research has concentrated on enhancing or modeling learner profiles based on interactions, behaviors, and personal information in e-learning environments for a variety of goals. As far as we are aware, little effort has been done to systematically build or enrich the learner profile with his social information and based on his activity and involvement within social media.

In the next chapter, we will present our contribution that consists of building a social profile ontology based on the learner's information within both MOOCs and social media environments.

Chapter 3: A social profile ontology to enhance the learning experience within MOOCs

Chapter 3: A social profile ontology to enhance the learning experience within MOOCs

1. **Introduction**:

Learner profile has emerged as a feasible model that support and promote the provision of learning opportunities. The use of learner profiles is one of the ways to adapt learning to the learners' specifics.
Modeling user profiles using publicly shared data on social media is referred as "Social Profiling" (Bilal et al. 2019). Social media has proved to be useful in many aspects to improve learning process and contribute to its success. Using social information and social behavior of users (clicks, posts, likes, etc.) for generating user profiles is beneficial since the User-Generated content hide the implicit interests of the user. Learners can be convinced and "seduced" into learning by providing them the support they need in finding the right courses that respond to their preferences.

The modeling of learners' profiles in MOOCs has been approached in different ways, considering their interactions, activities, and personal information within the e-learning environment. However, there have been limited attempts to extensively capture and enhance learner profiles by incorporating their characteristics, information, and interactions within social media platforms.

In this thesis, we implement an ontology-based profile approach by introducing our unique attributes and leveraging well-known ontologies and standards in order to implement our Social Profile Ontology (SPOnt) considering the learners' interests and preferences. In particular, we are extending concepts from IMS LIP, FOAF, Schema and DCMI standards, Felder-Silverman taxonomy, SIOC, Organization and Emoji ontology to represent learners' profile in both MOOCs and social media environments. We first create local ontologies and then, we rely on ontology mapping and merging techniques, to generate the global ontology that constitutes the Social Profile Ontology (SPOnt).

In the following sections, we highlight a detailed description of the different stages of the building of our Social Profile Ontology.

2. **Overall architecture of ontology building:**

The process of modeling a learner profile using ontologies encompasses various key aspects, including:

- Analyzing the learner 's requirements in MOOCs and Social Media environments (i.e., analyzing the learner 's personal and behavioral characteristics);
- Establishing the essential ideas that can describe and represent these learner characteristics and stating any potential features for these concepts;
- Creating the connections among the ontology concepts.
- Hierarchical classification of such notions and properties;
- Encoding these concepts and relationships using ontology building tools, and representing them in a formal ontology language for usage in adaptive applications.

Figure 7 depicts the overall architecture of building the social profile ontology. It consists on:

- Identifying the key concepts that constitutes the learner profile in both MOOCs and social media.
- Creating two local ontologies that represents the learner profile in MOOCs and social media respectively using Protégé.
- Relying on ontology mapping and merging techniques to generate the learner social profile using COMA 3.0 tool.

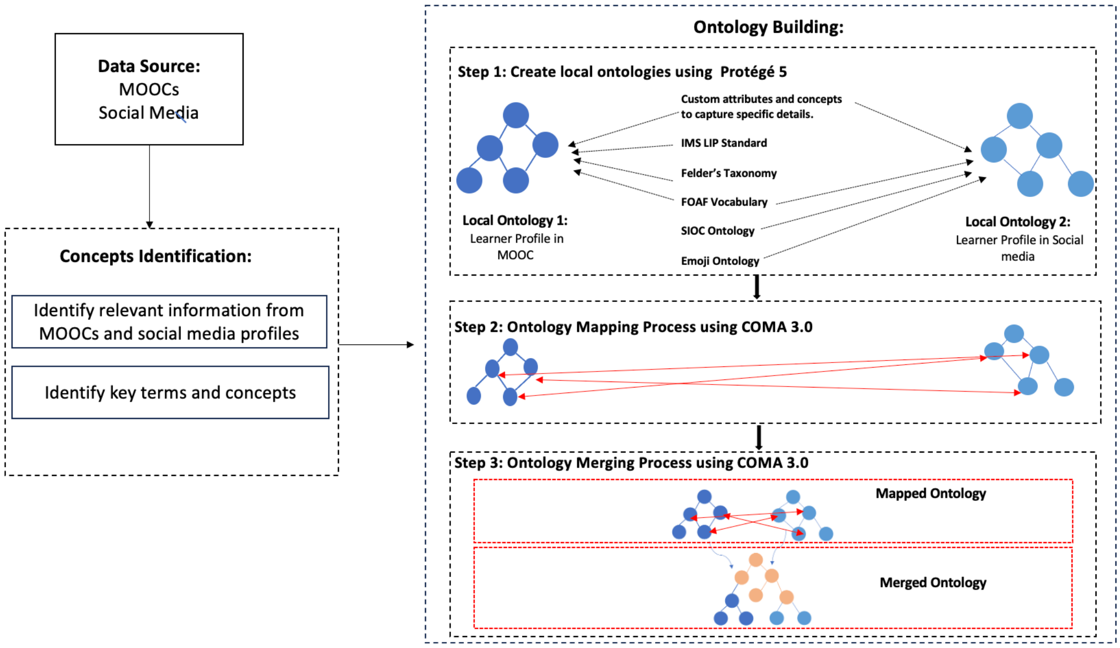

Figure 7 Overall architecture of Ontology Development

The creation of the two local ontologies has been achieved by including our custom attributes after investigating a set of profiles in MOOCs and social media. We reused, extended and integrated as well some standards and ontologies such as IMS LIP, FOAF, Felder's taxonomy, SIOC ontology, etc.
In the next section, we will highlight these concepts in details, as well as the process of ontology mapping and merging with COMA 3.0 tool.

2.1. Concepts:

The local ontologies have incorporated many terms from other standards and ontologies along with our custom concepts:

2.1.1. IML LIP standard:

("IMS Learner Information Packaging Information Model Specification | IMS Global Learning Consortium," n.d.): is employed as the gold standard for modeling and storing learner profile data: A data model describing the characteristics of learners serves as the foundation for the IMS Learner Information Package. Identity, QCL, Accessibility, Activity, Aim, Competence, Interest, Transcript, Affiliation, Security key, and Relationship are the eleven basic categories into which the learner's information is divided. In other words, just the necessary data needs to be kept and packed; these categories reflect the fundamental data structures needed to support learner information. (As depicted in Figure 8) The section on ontology implementation will go into further detail about the various components.

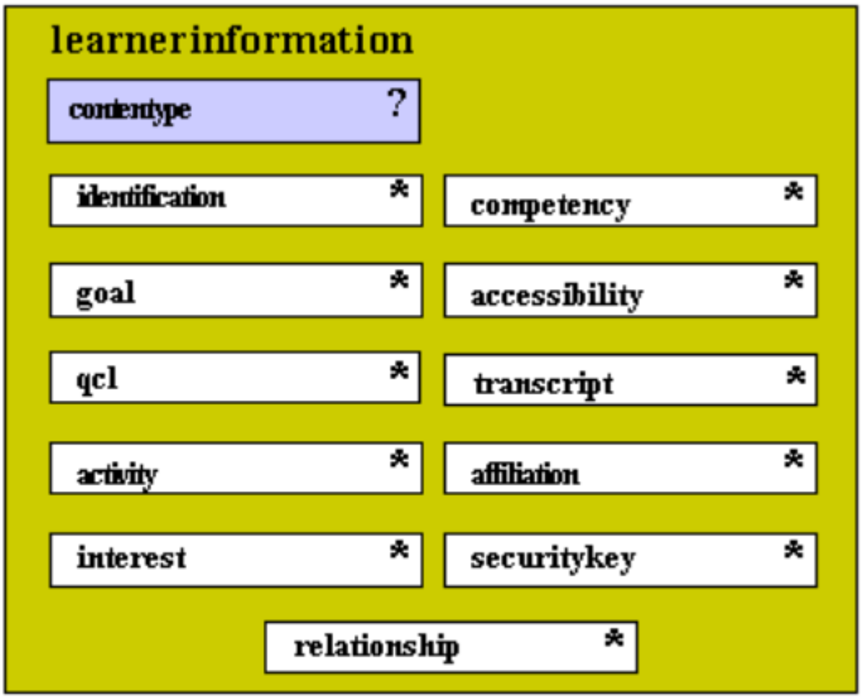

Figure 8 Learner Information Package (LIP) data

2.1.2. ***Friend Of A Friend (FOAF) vocabulary:***

FOAF project was established in 2000 to publish and exchange descriptions of people, their creations, and activities. The FOAF vocabulary standard provides a few fundamental types of entities (Person, Organization, Group, and Document), as well as the relationships that typically exist between them (Graves, Constabaris, and Brickley 2007). According to the "FOAF Vocabulary Standard", FOAF is a dictionary of terms that are either a class or a property ("FOAF Vocabulary Specification," n.d.). It describes the world using straightforward concepts drawn from the web. Some of the classes and properties that specifically characterize the social web were reused.

2.1.3. **Schema:**

With the goal of developing, preserving, and promoting schemas for structured data on the Internet, in web pages, email messages, and elsewhere, Schema.org is a cooperative community project. RDFa, Microdata, and JSON-LD are just a few of the several encodings that can be utilized with the Schema.org vocabulary. Through a well-documented extension model, these vocabularies—which include entities, connections between things, and actions—can be easily expanded (Iliadis et al. 2022).

2.1.4. *The Dublin Core Metadata Initiative (DCMI):*

RDF vocabularies are used to represent DCMI metadata terms so they can be used with Linked Data. By ignoring both the global identifier and the formal implications of the RDF-specific elements of term definitions, authors of non-RDF information can use the terms in settings like XML, JSON, UML, or relational databases. These users might focus on the definitions, usage notes, and examples' natural-language text and approve domain, range, sub-property, and subclass relations as usage suggestions (A. Zhang, Zhang, and Jia 2023).

2.1.5. *Felder-Silverman taxonomy*:

A number of different learning style models have been proposed in the literature, including those by Felder and Silverman (Felder 2002), Mayer and Myers (Briggs Myers and Myers 1995), and Kolb (El-Bishouty et al. 2019). Each of these models offers a different description and classification of the various learning types. In adaptive e-learning systems, learning style models are frequently employed to simulate how students learn. Learning styles are "the distinctive cognitive, affective, and physiological responses that serve as relatively stable indications of how learners perceive, interact with, and respond to the learning environment," according to Keefe (Keefe and Kiernan 1979). Learning styles offer knowledge about the various learning methods that students like, and as a result, they offer important details about student preferences. This knowledge can therefore be used to optimize the learner's learning process.

We are relying on the Felder-Silverman learning style framework in our work (FSLSM). Felder and Silverman explain a learner's learning style in greater detail by differentiating between preferences on four dimensions, in contrast to the majority of existing learning style models that divide learners into a small number of categories. is a four-dimensional model, according to Howard, Carver, and Lane (Howard, Carver, and Lane 1996):

- **Perceive Dimension (**“Sensing” / “Intuitive”): Identifies the preferred method of perception or knowledge absorption for the learners.
- **Process Dimension (**“Active” / “Reflective”): Specifies the preferred method of information processing by the learners.
- **Receive Dimension** (“Visual” / “Verbal”): Describes the preferred method of information presentation for the learners.
- **Understand Dimension** (“Sequential” / “Global”): Illustrates how learners prefer to arrange and advance through information.

2.1.6. *Organization ontology:*

This ontology has been created with the purpose of facilitating the dissemination of information concerning organizations and their structures, including governmental organizations. The main objective is to offer a versatile and adaptable core ontology that can be expanded or customized to suit specific contexts (“The Organization Ontology,” n.d.).

The ontology provides a set of terms to assist in representing the following:

- ***Organizational structure:***
  - Concept of an organization
  - Breakdown into sub-organizations and units
  - Purpose and categorization of organizations
- ***Reporting structure:***
  - Membership and reporting relationships within an organization
  - Roles, positions, and the connections between individuals and organizations
- ***Location details:***
  - Sites or buildings, as well as locations within those sites
  - Organizational history, including mergers and renaming events.

Figure 9 presents the main classes and relationships in the organization ontology:

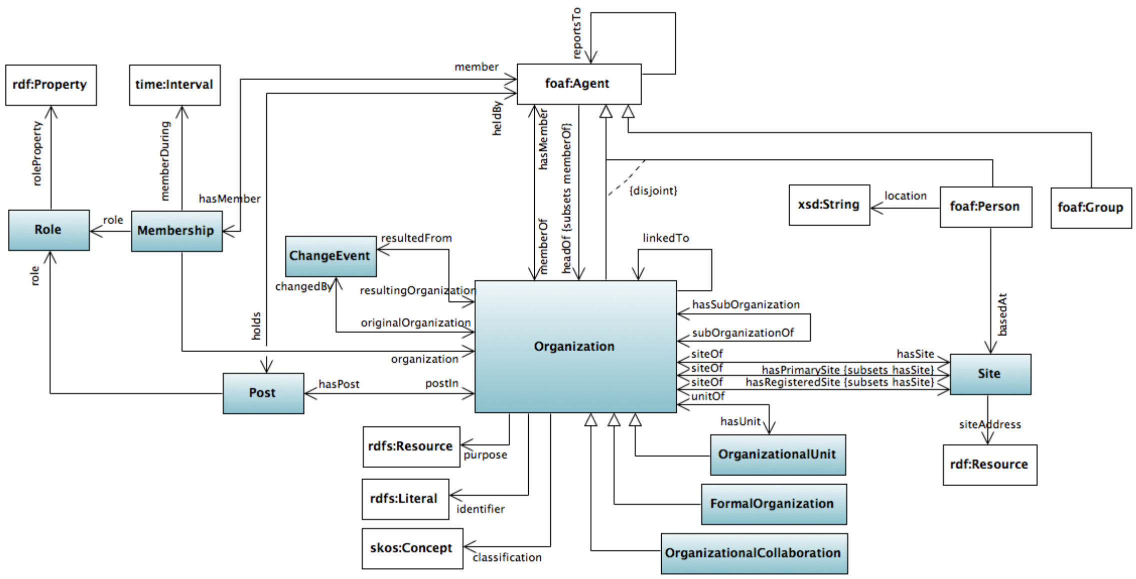

Figure 9 Overall illustration of Organization ontology

2.1.7. **SIOC ontology** ("SIOC Core Ontology Specification," n.d.):

The SIOC (Semantically-Interlinked Online Communities) Core Ontology offers the core concepts and attributes necessary to characterize data from online communities (such as message boards, wikis, weblogs, etc.) on the Semantic Web as presented in Figure 10. The SIOC Core Ontology offers more classes and properties in addition to those shown in this Figure.

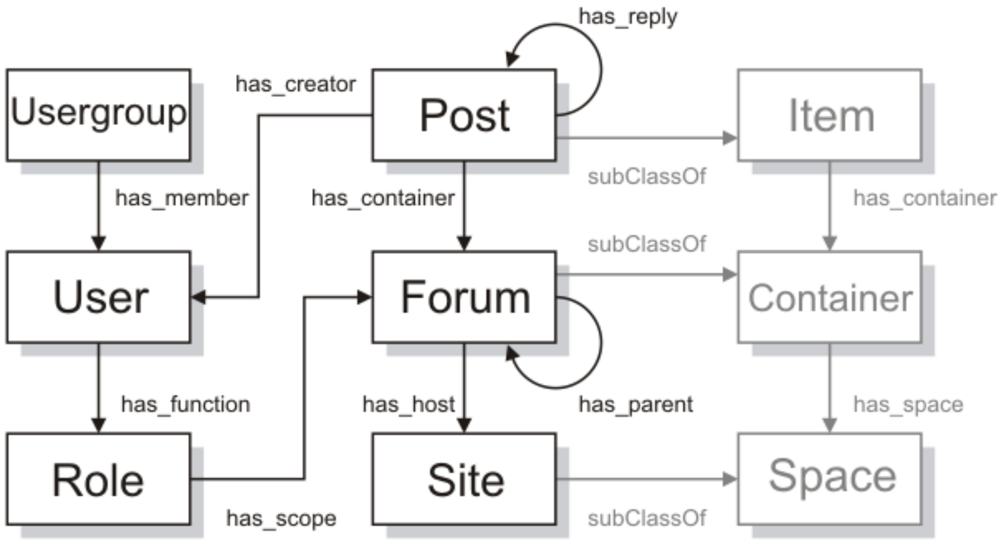

Figure 10 Main classes and properties in the SIOC ontology

2.2. Ontology mapping and merging process:

The two local ontologies (O1, O2) were mapped and then merged using COMA 3.0, which conducts matching and merging on relational database schemas, XSD (XML Schema), OWL (OWL-Lite), XDR (XML Data Reduced), and OWL (OWL) (Do and Rahm 2002).

The matching algorithm tool COMA 3.0 implements an iterative approach based on a variety of matching algorithms(matchers) for ontology mapping (Li et al. 2016). COMA 3.0 is divided in 4 modules as presented in Figure 11:

- **Storage**: This is mainly composed of the *Importers*, which upload schemas, ontologies, previous mappings, and auxiliary data (such instance data, dictionaries, etc.) into the repository where they are persistently kept. These files are directly accessible from the repository and can be utilized to complete related tasks.
- **The Match Execution:** is the core of COMA. It takes as input two schemas or ontologies, employs a number of matching methods, and computes the match outcome. The Execution Engine selects the pertinent schema components for matching in this module, uses a variety of matching techniques, and then combines the partial results to provide the whole match result. The acquired mapping can be fed into the following iteration for more fine-tuning.
- **The Mapping Processing module:** Once the match result is determined, more actions are carried out using the Mapping Processing module. It enables automatic enrichment of mappings (such as the detection of complex correspondences), model merging (ontology merging), and data transformation, either directly or by using a query script.

- **The User Connection module** consists of a full-fledged GUI, a SaaS-solution and an API.

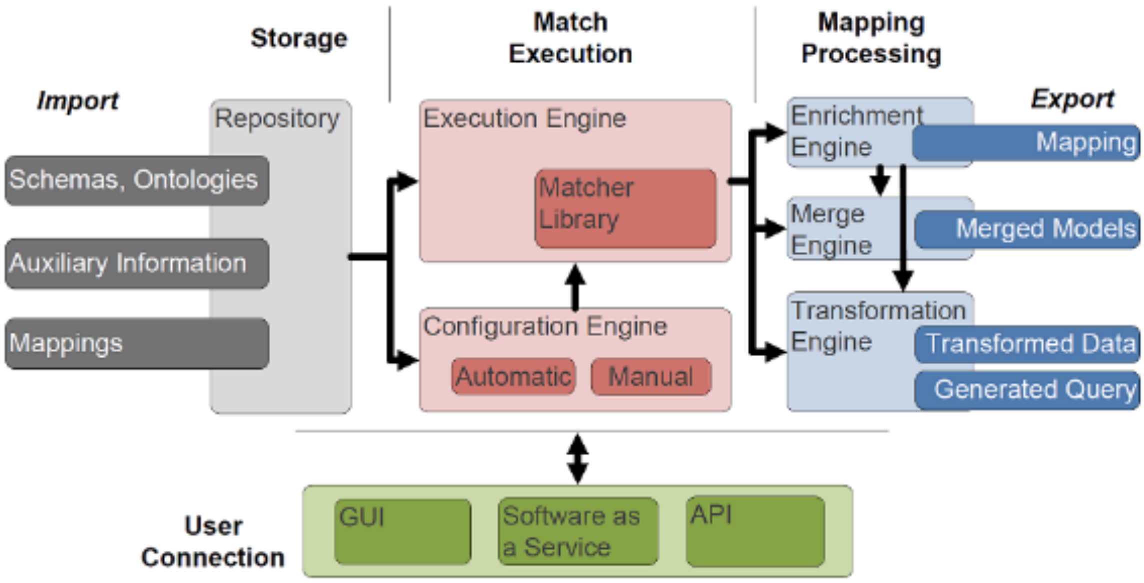

Figure 11 COMA 3.0 architecture

The ontology merging process is described as follow in COMA 3.0: The procedure uses the match mapping as input and generates a merged ontology, which is an integrated ontology. In our case, the input is the mapped ontologies and the output is the global ontology that contains our social profile.

The following section presents in details the building process of our social profile ontology using *METHONTOLOGY* methodology.

3. **Process of building the Social Profile Ontology:**

The social profile ontology was carried out through three major steps as presented in Figure 12 and which contains:

- Local ontologies creation process using METHONTOLOGY method.
- Ontology mapping process using COMA 3.0.
- Ontology merging process using COMA 3.0.

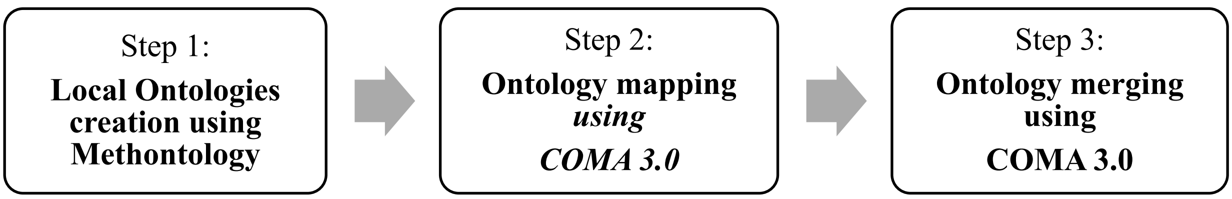

Figure 12 Sequences of steps for the implementation of the social profile ontology

3.1. Local ontologies creation process using METHONTOLOGY method:

METHONTOLOGY, one of the most well-known and complete methodology for ontology development, serves as the foundation for the methodological approach used to build, implement, and portray the ontology in our work.
METHODOLOGY is a well-structured approach that consists of a number of activities, techniques for carrying out each activity, and deliverables to be created after the completion of the activities using the techniques (Fernández-López, Gómez-Pérez, and Juristo 1997b). Figure 13 highlights the life cycle of building an ontology starting from planning and arranging the main tasks to be carried out, the time needed to perform this activity and the resources used (people, software and hardware). The other activities presented in the development process of the ontology will be discussed in details in the following section.

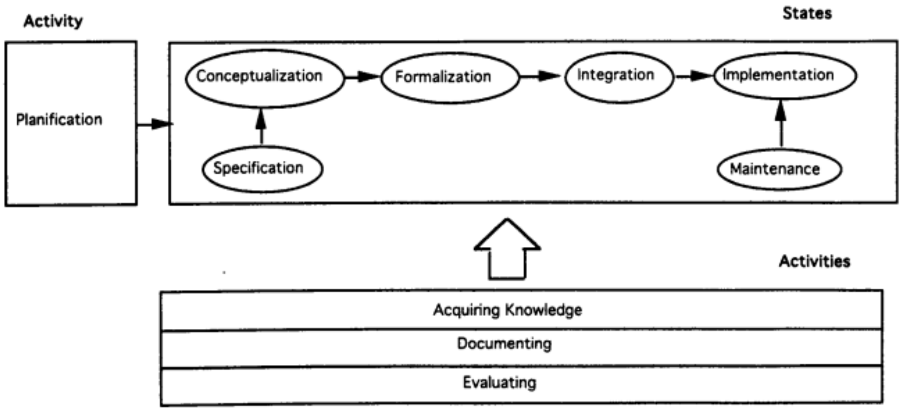

Figure 13 METHONTOLOGY life cycle

3.1.1. **Ontology specification:**

The ontology requirements specification document is what this phase intends to develop (ORSD). The document describes the main goal, function, level of granularity, and scope of the ontology (Yulianti and Surendro 2018). After a literature review of ontology- based user profile in the area of e-learning, we determined its ORSD, as presented in Table 10.

Table 10 ontology requirements specification document (ORSD)

Domain	E-learning, MOOCs and social media
Purpose	Build a social learner profile that contains information about leaners from his interaction and participation in both MOOCs and social media
Level of formality	Formal

Scope	List of the different concepts that model the user profile in both MOOCs and social media environments, including the interests and preferences of the user, personal involvement including behavior vis-à-vis a post (like, comment, share), learning styles, etc.
Requirements and competency questions	Requirements are analyzed based on study cases about the challenges that surround learners in their process of learning including drop-out, lack of satisfaction, lack of interaction.

3.1.2. **Knowledge acquisition:**

This phase's goal is to collect and elicit the domain knowledge needed to create the ontology (Waqialla et Razzak 2016). Our solution uses a literature review approach to gather information for modeling ontologies from specification through conceptualization about learners in MOOCs. Moreover, knowledge regarding user profiles has been gathered from a variety of sources, such as books, papers, well-known standards, and other ontologies that are already in existence. We began by defining the concepts, relationships, and properties for the primary classes in each ontology.

3.1.3. **Conceptualization:**

This phase's goal is to organize the domain knowledge into a conceptual model that expresses the issue and possible solutions in terms of the domain keywords during the ontology specification activity (Waqialla and Razzak 2016). The conceptualization activity entails the set of knowledge structuring tasks depicted in Figure 14:

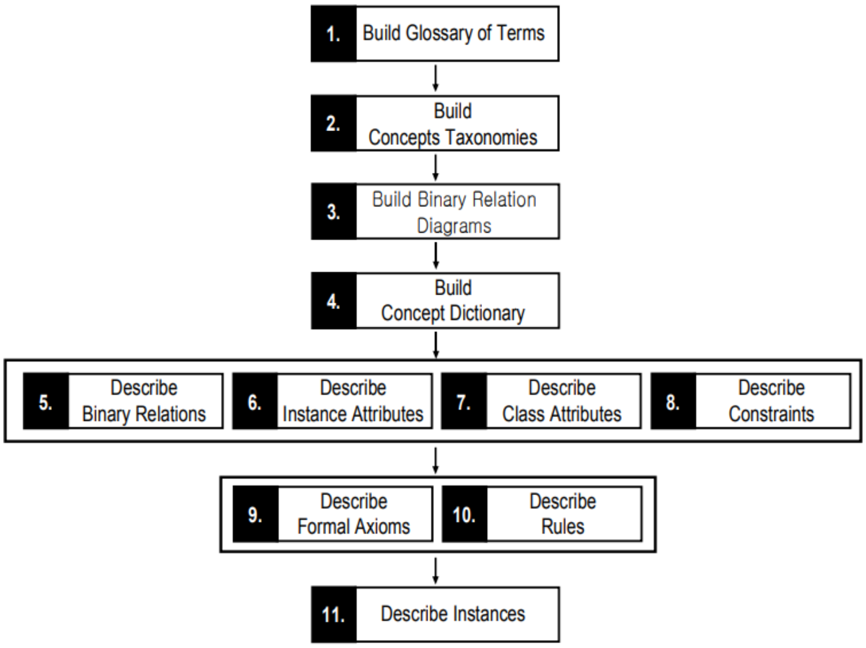

Figure 14 knowledge structure of conceptualization phase

3.1.3.1.***Task 1: Glossary of terms***

The glossary of terms task's objective is to compile a list of all the terms that are pertinent to the field (concepts, instances, attributes, relationships between concepts, etc.), together with their natural language explanations, synonyms, and acronyms.

Table 11 illustrates the glossary of terms for the local ontology that represents the learner profile in MOOC environment.

Table 11 An excerpt of the glossary of terms of ontology-based learner profile

Name	Synonyms	Description	Type
Learner Profile	Profile	Describes learner's characteristics and information	Concept
Personal Information	Basic Information Identification	Personal information such as name, address contact info, agent and demographics.	Concept
Course Preferences	Preferences Learning Preferences	Describes the learner preferred way for learning including: Preferred learning resources and learning styles	Concept
Education	--	Describes the learner 's education information including: university, degree, field of study, etc.	Concept

Social media accounts	User account Online account	Describes the individual social media accounts like Gmail, Facebook, LinkedIn, Twitter, Instagram, etc.	Concept
Has Profile	--	--	Relation
Has highest degree	--	--	Relation
Holds account	--	--	Relation
Account Name	--	Describes the name of the online account	Attribute
etc	etc	etc	etc

After highlighting the glossary of terms for the learner profile ontology in MOOC environment, Table 12 presents an excerpt of the glossary of term of the ontology-based learner profile in social media.

Table 12 An excerpt of the glossary of terms of ontology-based user profile

Name	Synonyms	Description	Type
Privacy setting	--	Relates to learner's decision regarding his profile to be public or private	Concept
Likes	--	Describes the type of likes that a post receives. It can be a smiley emoji, crying face, a heart emoji, etc.	Concept
Groups	User group	Represents the groups the user is a part of on social media	Concept
Notification	--	Represents the notifications the user receives on social media	Concept
Notification type	--	Represents the type of notification either it is a 'Comment Notification 'or 'Like Notification 'or 'Profile Visit Notification ', etc.	Concept
Has follower	--	--	Relation
Has profile photo	--	--	Relation
Received notification	--	--	Relation
Gender	--	--	Attribute
etc	etc	etc	Etc

3.1.3.2.*Task 2: Concept Taxonomies*

METHONTOLOGY suggests using the four taxonomic relations known as Subclass Of, Disjoint-Decomposition, Exhaustive-Decomposition, and Partition that were specified in the Frame Ontology (Farquhar, Fikes, and Rice 1997) and the OKBC Ontology (Corcho, Fernández-López, and Gómez-Pérez 2006).

- ***Subclass Of:*** A concept C1 is a Subclass-Of another concept C2 if and only if each instance of C1 is likewise an instance of C2. A concept can be a subclass of more than one concept in the taxonomy.

- ***Disjoint-Decomposition of a concept C*** is a set of subclasses of C that do not share instances and do not cover C, meaning that instances of the concept C may exist that are not instances of any of the concepts in the decomposition.

- ***An Exhaustive-Decomposition of a concept C*** is a collection of subclasses of C that include C and may have shared instances and subclasses; hence, there cannot be instances of C that are not also instances of at least one of the concepts in the decomposition.
- ***A Partition of a concept C*** is a set of subclasses of C that do not share common instances and that cover C, that is, there are not instances of C that are not instances of one of the concepts in the partition.

3.1.3.3.**Task 3:** ***Ad hoc binary relations diagram***

Ad hoc binary relations diagrams are used to create ad hoc connections among concepts belonging to the same (or other) concept taxonomy (Corcho et al. 2005). Figure 15 presents a fragment of the ad hoc binary relation diagram of the local ontology-based learner profile in MOOC.

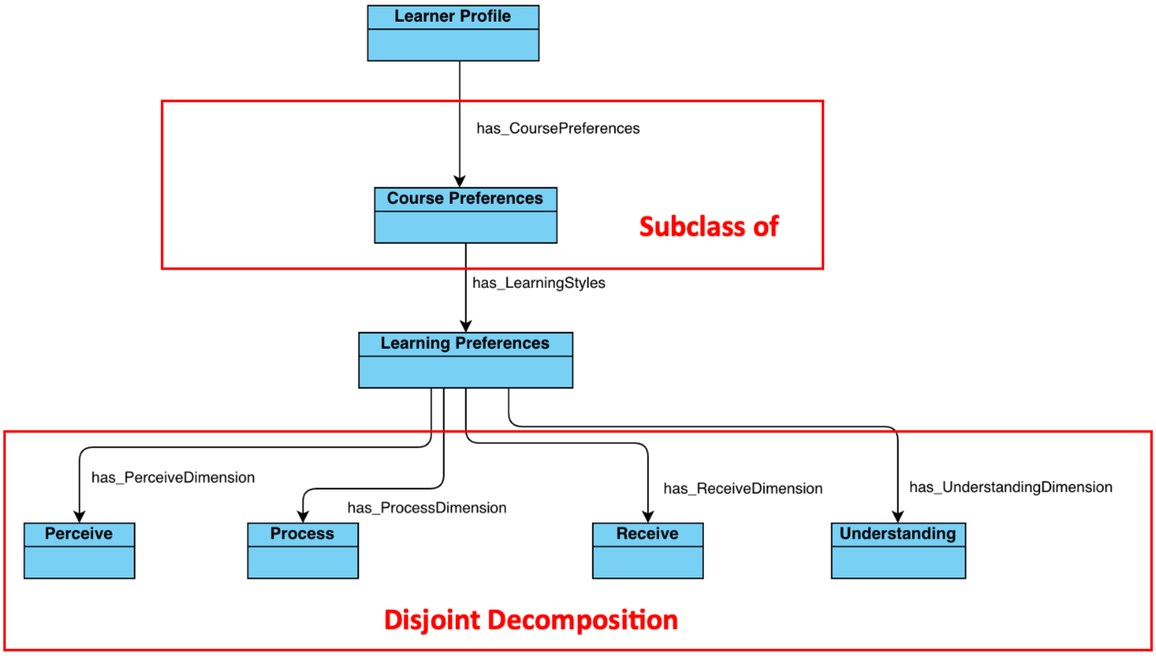

Figure 15 ad hoc binary relation diagram of learner profile ontology

Figure 16 illustrates the ad hoc binary relation diagram of the local ontology-based learner profile ontology in social media.

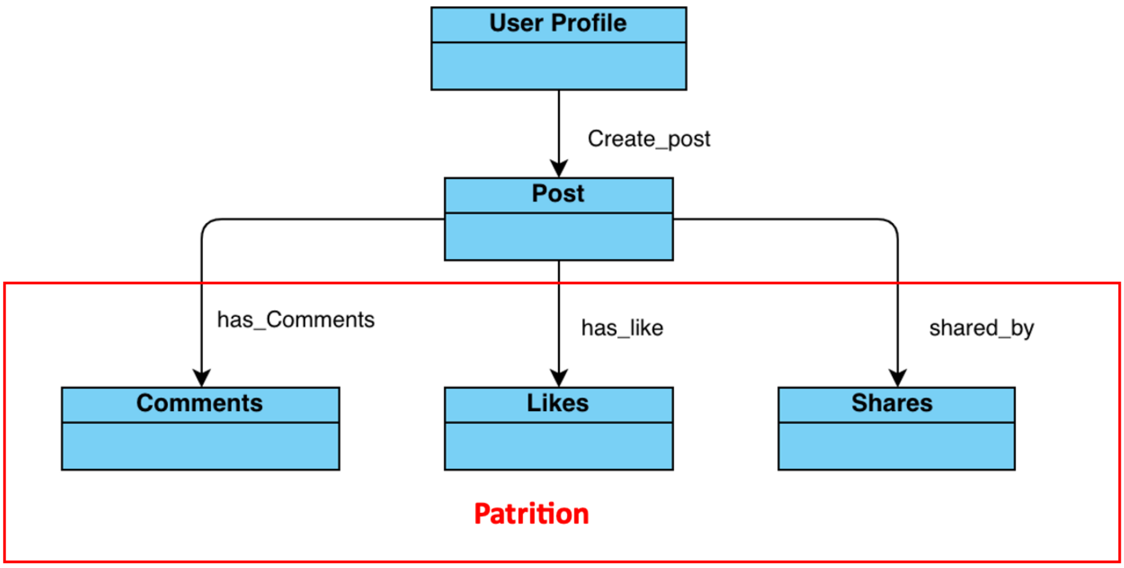

Figure 16 ad hoc binary relation of user profile ontology

3.1.3.4.*Task 4 : Concept dictionary*

Each domain concept, along with all of its relations, instances, and class and instance characteristics, is contained in a ***concept dictionary*** (Corcho et al. 2005). Table 13 highlights an excerpt of the concept dictionary of ontology-based learner profile in MOOC.

Table 13 An excerpt of the concept dictionary of ontology-based learner profile

Concept name	Instance	Class attribute	Instance attribute	Relations
Learner profile	---	---	---	Has_PersonalInformation Enrolled_in Has_CoursePreferences Has_WorkExperience Has_skill Holds_account etc.
Course Preferences		---	Course Name Duration Skills	
Learning Styles	---	---	Learning Style Type Learning Style name	Has_LearningStyle
Learning Resources	PDF Videos Images etc.	---	Learning resource type Learning resource format	Has_LearningResources
etc	etc	etc	etc	Etc

Table 14 presents an overview of the concept dictionary for the ontology-based learner profile in social media.

Table 14 An excerpt of the concept dictionary of ontology-based user profile

Concept name	Instance	Class attribute	Instance attribute	Relations
User Profile	---	---	---	Has_follower has_PersonalInformation Holds_account Has_PrivacySetting Friend_of Has_interest etc.
Online Account	Facebook Instagram LinkedIn Twitter	--- ---	Account URL	Account_of Holds_account etc.

	etc.			
User Account	--	--	Account name	Account_of Friend_of Has_Follower Has_PrivacySetting etc.
Post	---	---	Creator Content Post date Post likes count Post shares count	Has_post shared_by created_by etc.
Group	---	---	Group_name Group_description moderator_name	Has member Has moderator etc.
etc	etc	etc	etc	etc

*3.1.3.5.**Task 5: ad hoc binary relation in details***

This task's objectives are to construct the ad hoc binary relation table and to thoroughly characterize each ad hoc binary relation found in the concept dictionary.

We list the name, the names of the source and target concepts, the cardinality, and the inverse relation for each ad hoc binary relation.

Tables 15 shows a section of the ad hoc binary relation table of the ontology-based learner profile in MOOC.

Table 15 An excerpt of ad hoc binary relation for ontology-based learner profile

Relation Name	Source Concept	Source Cardinality (Max)	Target Concept	Inverse Relation
Has Personal Information	Learner Profile	1	Personal information	--
Has Courses Preferences	Learner profile	N	Course Preferences	--
Has Learning Styles	Course Preferences	N	Learning Styles	--
etc	etc	etc	etc	etc

Table 16 illustrates an extract of the ad hoc binary relation regarding the ontology-based learner profile in social media.

Table 16 An excerpt of ad hoc binary relation for ontology-based user profile

Relation Name	Source Concept	Source Cardinality (Max)	Target Concept	Inverse Relation
Holds_account	User Profile	N	User account	--
Has Personal Information	User profile	1	Personal Information	--
Has attachment	Post	N	Attachment	--
etc	Etc	etc	etc	Etc

3.1.3.6. **Task 6 : Instance attributes in details**

The purpose of this task is to create an instance attribute table that uses detailed descriptions of all the instance characteristics that are currently present in the concept dictionary. The instance attribute table's row-by-row explanation of each instance attribute is contained within each row. A concept's instances are described via its instance attributes, which may have a distinct value for each instance. We list the following information for each instance attribute: its name, the concept it belongs to (attributes are specific to concepts), its value type, and range of values (in the case of numerical values); minimum and maximum cardinality; instance attributes, class attributes, and constants used to infer values of the attribute; attributes that can be inferred using values of this attribute; formulas or rules that allow inferring values of the attribute; and references used. Table 17 presents a fragment of the instance attribute table of ontology-based learner profile.

Table 17 An excerpt of attributes for ontology-based learner profile

Instance Attribute Name	Concept Name	Value Type	Value Range	Cardinality
Name	Personal Information	String	---	(1,1)
Gender	Personal Information	String	---	(1,1)
Learning Style Type	Learning Styles	String	---	(1,1)
etc.	etc.	etc.	etc.	etc.

Table 18 shows an excerpt of attributes for ontology-based learner profile in social media.

Table 18 An excerpt of attributes for ontology-based learner profile

Instance Attribute Name	Concept Name	Value Type	Value Range	Cardinality
Likes count	Likes	Integer	---	(1,1)
Comment content	Comments	Literal	---	(1,1)
Account name	User Account	String	---	(1,1)
etc.	etc.	etc.	etc.	etc.

3.1.3.7.**Tasks 7,8,9,10 &11:**

These activities are designed to thoroughly characterize all of the class properties, constants, formal axioms, and rules that are already listed in the concept dictionary. After the ontology's conceptual model has been developed, it is crucial to specify pertinent instances that exist in the concept dictionary inside an instance table. (Corcho et al. 2005).
Figures 18 and 19 presents an overview of the two local ontologies including their classes and subclasses.

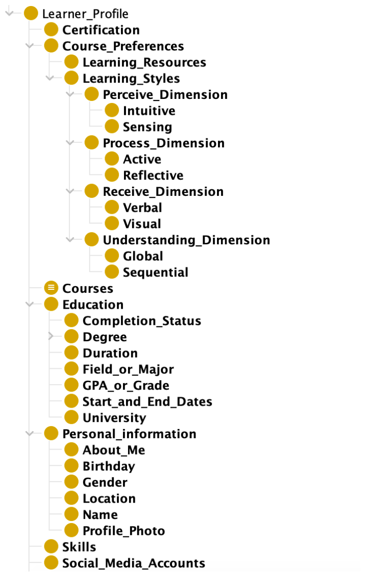

Figure 19 Overview of ontology-based learner profile concepts

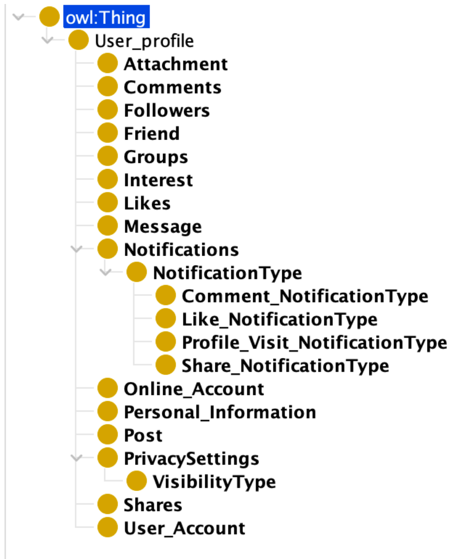

Figure 18 overview of ontology-based user profile concepts

Figures 20 and 21 show an excerpt of object properties of the two local ontologies:

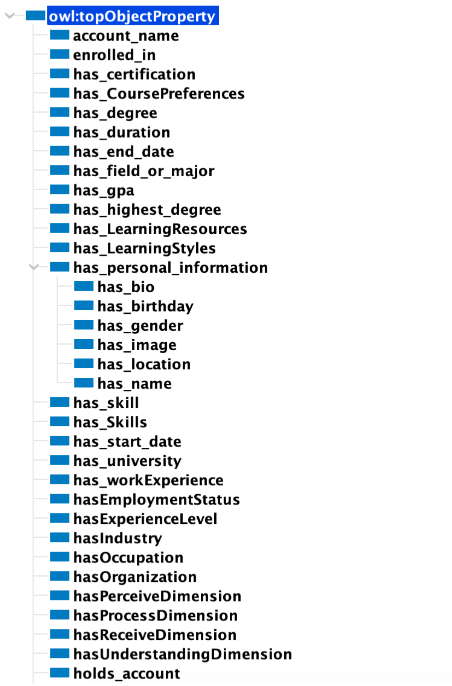

Figure 21 Excerpt of the object properties of the ontology-based learner profile

Figure 20 Excerpt of the object properties of the ontology-based user profile

3.1.4. Integration:

Throughout ontology development, we identified terms that could be included from other ontologies. we use well known standards for modeling the learner profile in MOOCs including: FOAF, IMS LIP standard as well as Felder’s taxonomy for describing the learning styles. For representing the user profile in social media, we use SIOC ontology and Emoji ontology.

3.1.5. Implementation:

The use of an environment that supports the meta-ontology and ontologies chosen during the integration phase is necessary for the deployment of ontologies. This phase ends with the ontology being formalized in a formal language. We used Protégé v5 to implement the two local ontologies.

Protégé (Kapoor and Sharma 2010) is an ontology and knowledge base editor created at Stanford University. Protégé is free, developed in Java, it can be extended, and a plug-and-play environment is provided, making it a versatile platform for application development and rapid prototyping. Protégé enables the definition of classes, class hierarchies, variables, variable-value limitations, relationships between classes, and the characteristics of these relationships. The most recent RDF and OWL 2 Web Ontology Language specifications from the World Wide Web Consortium are fully supported by Protégé. Protégé's strength is that it supports tool developers, knowledge engineers, and domain experts all at once.

Figure 22 represents the implemented ontology for the learner profile.

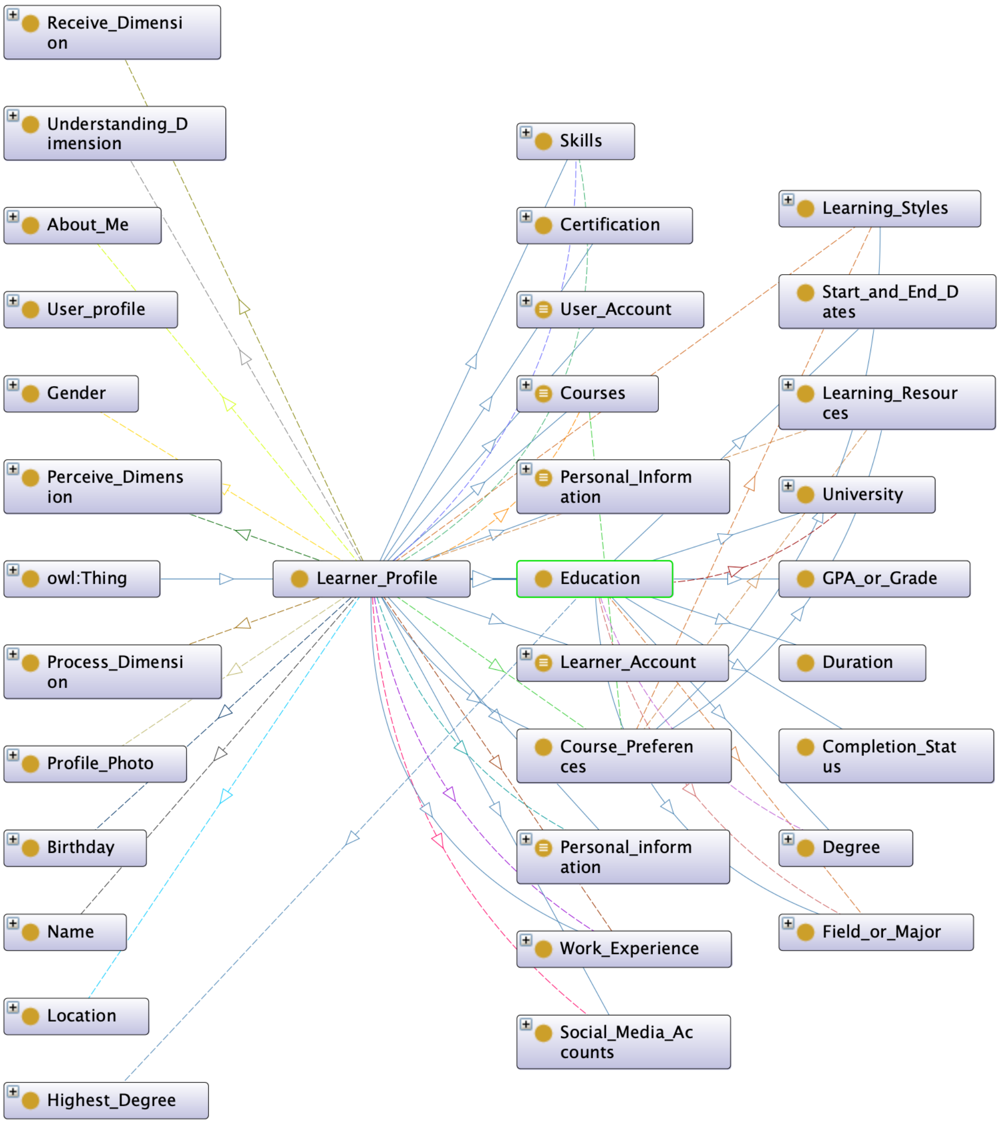

Figure 22 Graphical representation of ontology-based learner profile

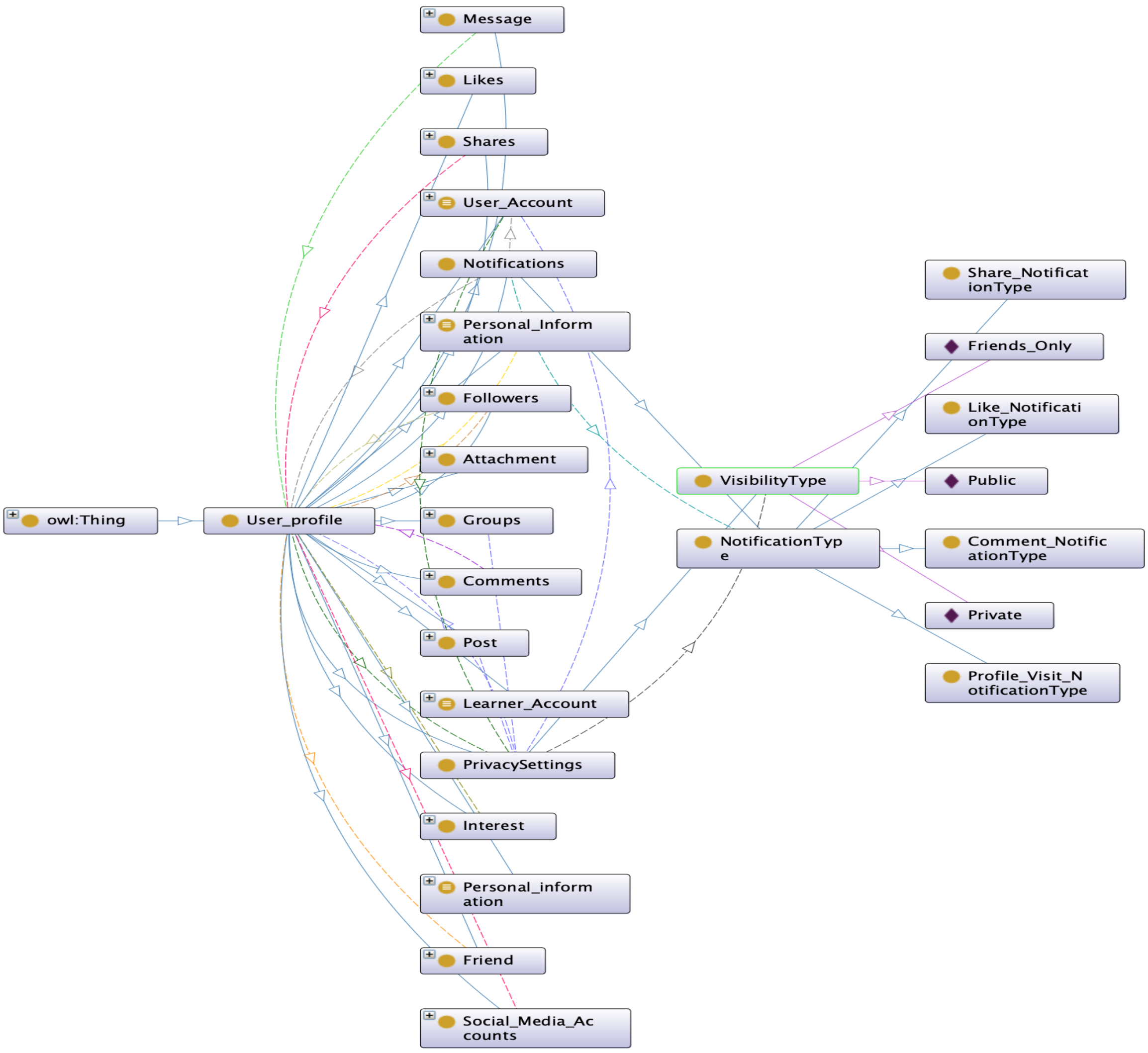

Figure 23 Graphical representation of ontology-based user profile

3.1.6. Evaluation:

Evaluation is the process of giving the ontologies, their software environment, and their documentation a technical assessment in relation to a frame of reference (in this case, the requirements specification document) both within and between phases of their life cycle. c. Verification is the technological process that ensures an ontology, the software environments that it is associated with, and the documentation are accurate with regard to a frame of reference at each stage and in between stages of their life cycle. According , validation ensures that ontologies, software environments, and documentation match the systems they are meant to represent.

The method of evaluation based on criteria has been chosen. With this method, SPOnt's usability is being evaluated. Our implemented ontology was improved as part of the criteria-based review in response to input from the experts in the field. These criteria are:

- ***Clarity:*** We conform to Gruber (Gruber 1995)’ specifications. In our case, we rely on the comments provided by the e-learning domain experts who were contacted to improve and confirm the ontology's clarity. For instance, we have "Learning Resources" and "Learning Styles" as two distinct concepts at the level of the local ontology that reflects the learner profile in MOOC. The both concepts reflect the preferences of the learner regarding the courses he wants to enroll in. In order to convey a more intended meaning, professionals in the field of e-learning prefer to create another concept called ‘Course Preferences ‘that includes both "Learning Resources” and "Learning Styles" as sub-classes of this main class.

- ***Consistency/Coherence****:* Based on the recommendations of experts, the consistency check produced a lot of examples where adjustments were suggested. Ontology notions ought to be logically consistent and free of ambiguity or inconsistencies.

- ***Conciseness****:* According to Gómez-Pérez (Gómez-Pérez, Fernández-López, and Corcho 2004), "ontology is concise if it does not include redundant or useless definitions aside from clear redundancies between definitions." For instance, the SPOnt ontology had the classes "Learner" and "learner Profile" provided. The second class is related to the notions that constitute his profile such as learning style, preferred learning resources, etc. The first class represents everything related to learners including personal information, degrees, and relationships. We added its sub-classes to the "learner profile" class after realizing that the class "learner" is unneeded.

- ***Correctness:*** "It is the correspondence between ontology concepts and properties and real-world entities and properties." In the creation and validation of our ontology, we gave significant thought to this requirement. Feedback from domain experts has been extremely helpful in ensuring that the ontology is accurate and verified.

3.2. Ontology mapping process using COMA 3.0:

In order to find similarities between the entities of the implemented ontologies, we use COMA 3.0. We applied three individual matchers: “AllContextW”, “NodesPathW” and “NodesNamesW”. Figure 24 below represents an overall architecture of the mapping process.

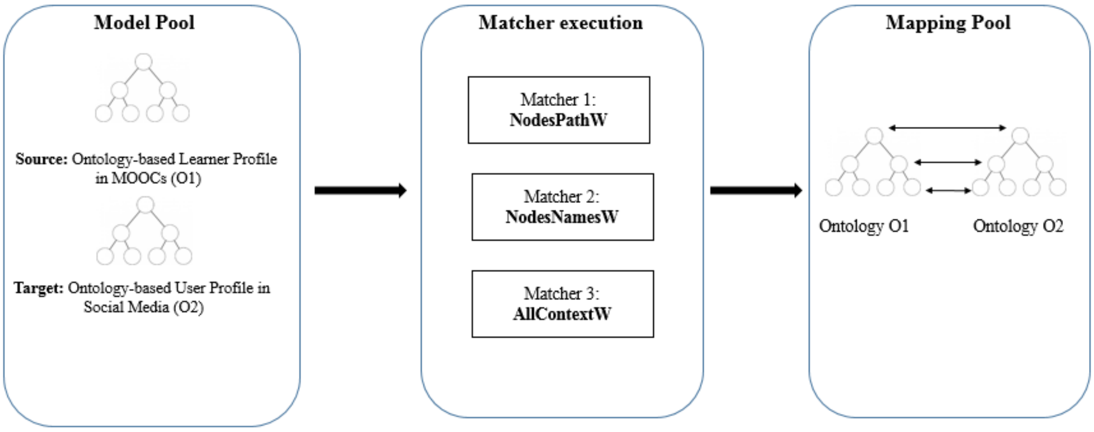

Figure 24 Overall architecture of the process of mapping

For each matcher, we present the matching results. Each line represents a correspondence between two concepts. In our case, we choose to point out the mapping between ‘account_name‘ concept in both ontologies. Figure 25 highlights the match result of ‘ALLContextW ‘matcher.

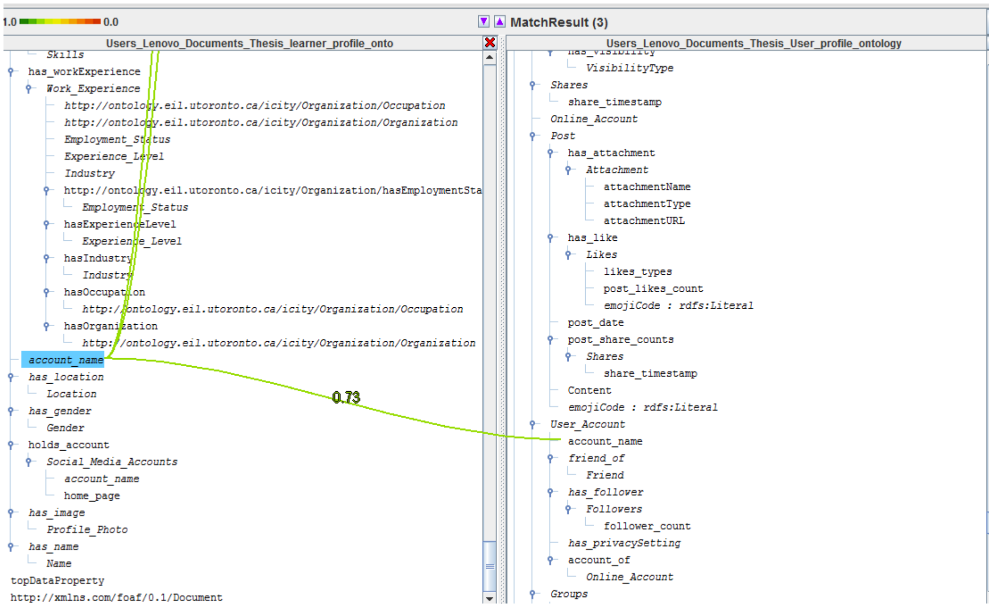

Figure 25 Match result using ALLContextW matcher

Figure 26 presents the match result of ‘NodesNameW ‘matcher.

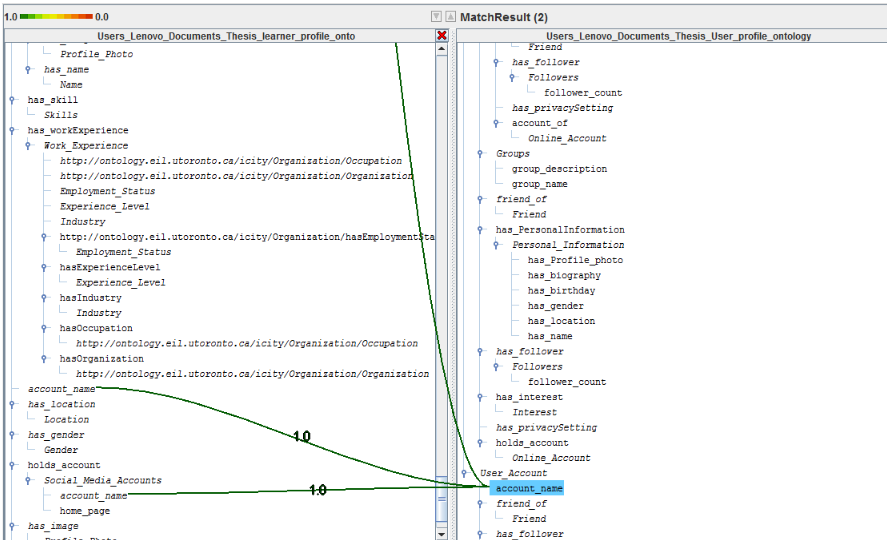

Figure 26 Match result using NodesNameW matcher

While figure 27, presents the match result of ‘NodesPathW‘matcher.

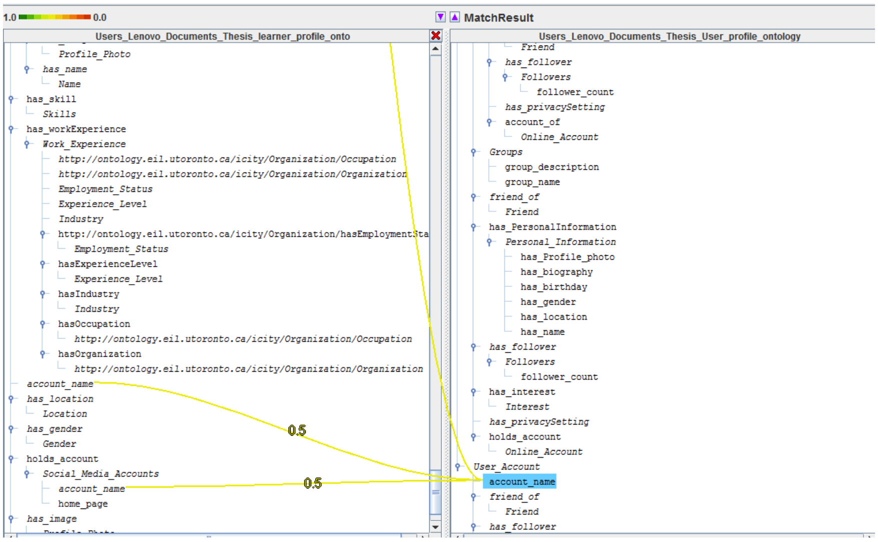

Figure 27 Match result using NodesPathW matcher

We choose to apply the “NodesPathW” matcher for the ontology mapping process as it gives more accurate match result. Figure 28, highlights the match result using “NodesPathW matcher”

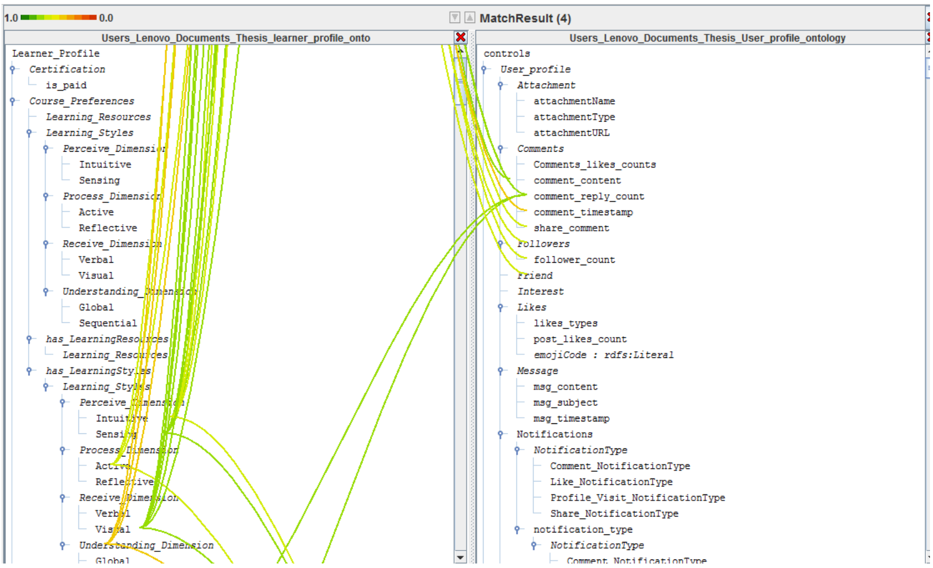

Figure 28 Match result using NodesPathW matcher

3.3. Ontology merging process using COMA 3.0:

Ontology merging is a process of combining two, or more, ontologies into one. Consequently, the resulting ontology stores knowledge from all merged ones. Merging often utilizes a set of alignments to create deep interconnections between ontologies and, in the end, merge them into one. We used also COMA 3.0 for the process of merging.
In order to evaluate the quality of the different combinations of matchers and to choose the best one to apply for our merging process, we use three metrics for this purpose: Recall, precision and F-measures.

Definition: Recall, Precision and F-measure

- ***Precision*** and ***Recall*** are commonplace measures in information retrieval. They are based on the comparison of an expected result and the effective result of the evaluated system (Makhoul et al. 1999).
  Given a reference alignment R: Precision and recall of some alignment A are defined by:
  $$\mathbf{P\ (A, R) = |R \cap A| / |A|}$$
  $$\mathbf{R\ (A, R) = |R \cap A| / |R|}$$
- ***F-measure*** is defined as a weighted combination of P and R.

We test the different combination based on the match result in order to obtain the best merge result. Figure 29 presents the result of the evaluation of the different combination

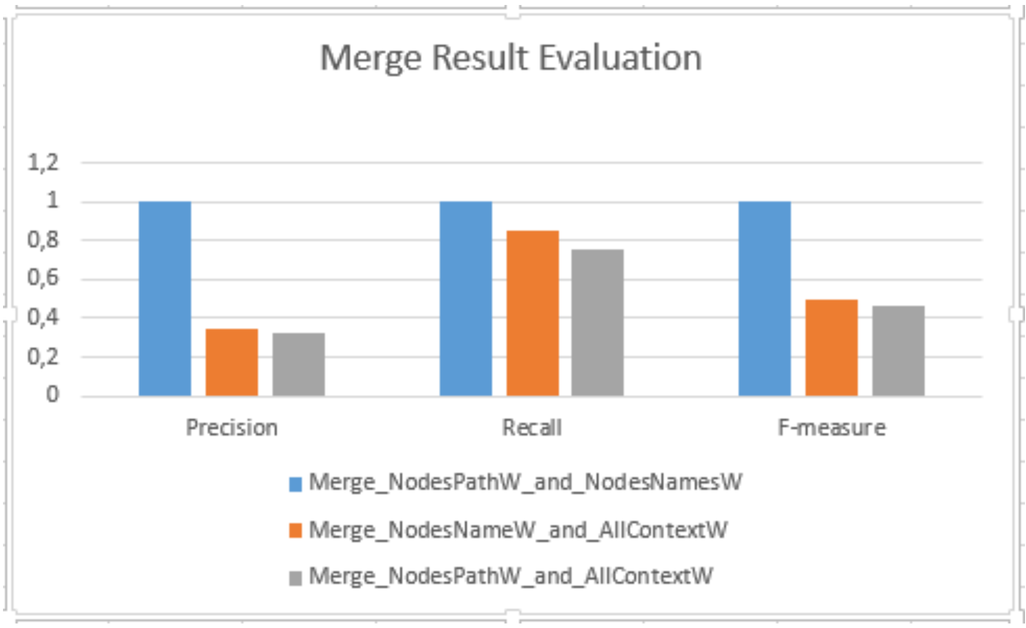

Figure 29 Evaluation Result of different combination between the match results

Based on the above evaluation, we merged the match result of both "NodesPathW" matcher and "NodesNameW" matcher. Figure 30 represents the result of the merge.

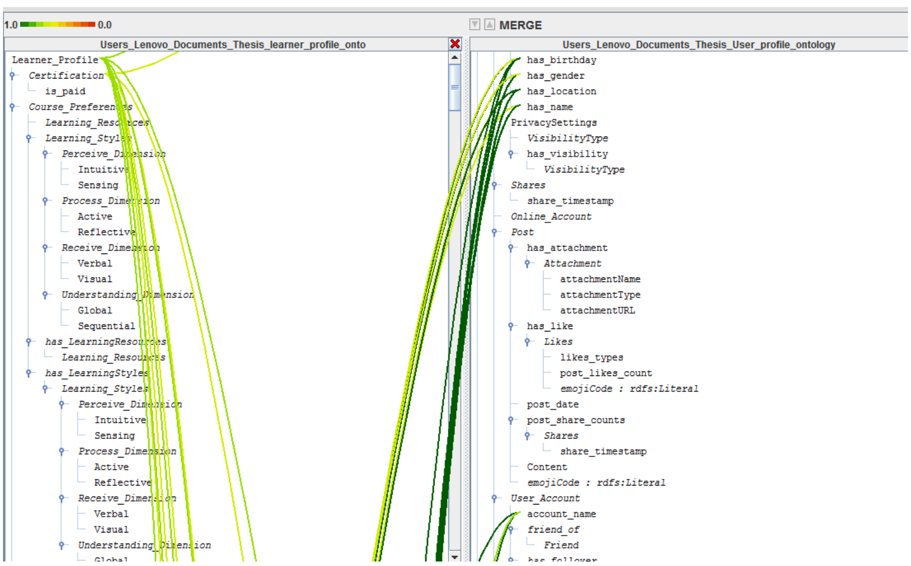

Figure 30 Merge Result of the merged ontologies

In conclusion, this thesis presents an innovative approach to modeling learners' profiles by leveraging the power of ontologies and standards to create a Social Profile Ontology (SPOnt). The goal is to capture learners' interests and preferences not only within the MOOC (Massive Open Online Course) environment but also by incorporating their characteristics, activities, and interactions from social media platforms.

The proposed approach involves the creation of two local ontologies—one for the learner profile in MOOCs and another for the user profile in social media. These local ontologies are built using our custom attributes and well-known standards and ontologies such as IMS LIP, FOAF, Schema.org, DCMI, Felder-Silverman taxonomy, and Organization ontology. Each of these contributes to capturing specific aspects of the learner's profile in MOOC and social media. The ontology mapping process is then employed to ensure interoperability between the different concepts in the local ontologies. COMA 3.0, a matching algorithm tool, is used for ontology mapping, enabling the generation of a unified view of the learner's social profile. Finally, the ontology merging process combines the mapped ontologies to create the Global Ontology (SPOnt), which represents the integrated learner social profile. This ontology can serve as a tool to provide personalized learning opportunities to learners by identifying their preferences and interests and guiding them to the right courses and resources.

We are more interested in the “Interest” component in our profile since interests play an important and vital role in promoting engagement and fulfilling the learners’ requirements. In the next chapter, we highlight the use of topic modeling and NLP techniques to extract the preferences of learners from their generated content on social media.

Chapter 4: Topic modeling for the detection of learner's interest based on the social profile ontology

1. Introduction:

The learners' interest forms the essential characteristics of the learner profile. In the context of MOOCs, Interests play an essential role in the learning process, they are a powerful motivational process, that energize learning, guide academic and career trajectories (McIntyre, Gundlach, and Graziano 2021) (Harackiewicz, Smith, and Priniski 2016). A learner interest is a key component of adaptive hypermedia and educational systems that focus on the learner's behaviors and personalize courses according to learner interests (Peng et al. 2016). Learners are motivated to invest time and effort toward the course of their interest, thereby, enriching learners' interest will yield to a better discovery of courses' subject that are a best fit with their preferences which impact their satisfaction and thereby their interaction inside MOOCs.

MOOCs already recommend courses that respond to learners' interest based on their participation. However, the same learners are more interactive in social media through the content they generate and which contains hidden information about their "real" interest and preferences. Since generating the user interest is a challenging task, using topic modeling techniques are useful to uncover the main thematic information related to user (Bai et al. 2021). We propose a Course Topic Model (CTM) based on Natural Language Processing (NLP) and topic modeling techniques to identify and detect learners' course of interest based on their spontaneous interaction in social media, in particular Twitter. The generated CTM contains the most probable topic of interest for each learner. We train three well known models: Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA) and BERTopic after applying NLP pipeline on the tweets shared by the learners. Then, we evaluate the models using topic coherence score, topic diversity score and human judgement. Experiments performed reveal that BERTopic and LDA model performed better on the scrapped dataset and their results are used to generate the course topic model.

This chapter, is divided into 2 main sections: the first one is dedicated to the background of our work and second one highlights our contribution in detail.

2. Background and definitions:

In this section, we point out the used techniques for our solution, notably, NLP and Topic modeling. We highlight their definitions and some of their tasks and algorithms.

2.1. Natural Language Processing (NLP):

2.1.1. NLP definition:

In the fields of computer science and artificial intelligence, NLP is of greatest priority. NLP research includes theories and techniques that facilitate effective natural language interaction between people and computers. According to the official definition, NLP is "a theoretically motivated range of computational techniques for studying and representing naturally occurring texts (of any mode or type) at one or more levels of linguistic analysis for the purpose of attaining language that is like a human-like language processing for a range of tasks or applications."(Suleman and Korkontzelos 2021).

NLP is a scientific discipline that combines computer science, linguistics, and mathematics. Its main objective is to convert human (or natural) language into computer commands(Kang et al. 2020). The field of NLP, also referred to as computational linguistics, entails the development of computer models and procedures to address real-world issues with language comprehension utilizing machine learning, statistical, and probabilistic calculations (Otter, Medina, and Kalita 2021).

The application domains cover subjects including information extraction (e.g., named entities and relations), text translation between and among languages, summarizing written works, automatically determining responses to queries by inferring answers, and categorization and clustering of documents. Frequently, one must successfully address one or more of the fundamental challenges before using the concepts and methods developed to address actual concerns.

2.1.2. NLP tasks:

NLP field encompasses a range of tasks, including but not limited to text summarization, Part of speech tagging, tokenization, lemmatization, removing stop words, etc.

- **Text summarization:** Involves using software to shortening a text document with software, in order to create a summary with the main ideas of the original document (Awasthi et al. 2021). Text Summarization methods are broadly categorized into different types as shown in the Figure 31 below:

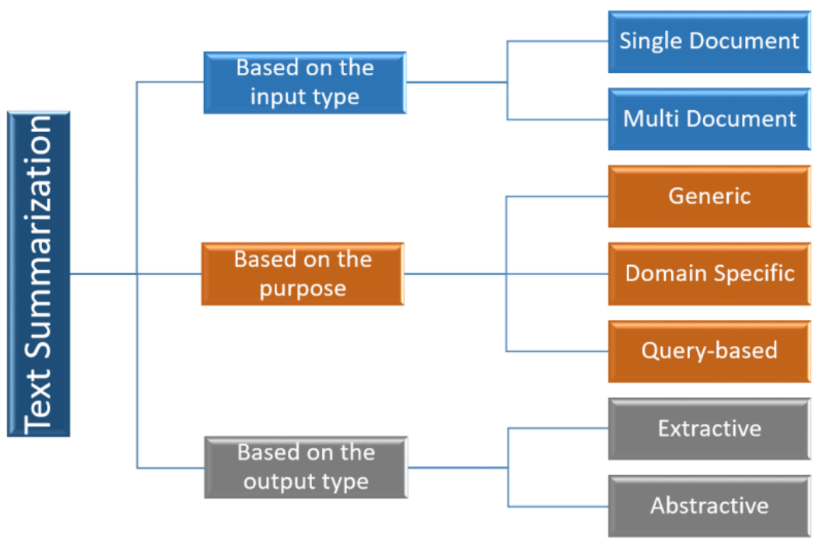

Figure 31 Text summarization methods (Boorugu et Ramesh 2020)

Text summarizers can be sorted by (Boorugu and Ramesh 2020):

- ***Input type:*** Single Document for short text and basic models and multi-document for longer text with more complexity.

- ***Purpose Type:*** summarizers are Generic (impartial treatment), Domain-specific (using domain knowledge), and Query-based (including answers to questions).
- ***Output type:*** Summarizers are Extractive (selecting important sentences) and Abstractive (creating coherent summaries like humans).

- **Part of speech tagging** (Voutilainen 2003),(Thavareesan and Mahesan 2020) also called grammatical tagging. It is the process of identifying the part of speech of a certain word or passage of text based on its use and context. Part of speech identifies 'make' as a verb in 'I can make a paper plane,' and as a noun in 'What make of car do you own?'.

- **Tokenization**: One of the most essential NLP tasks is tokenization. It involves dividing a text document into tokens, which are relatively short pieces of text. A token can be an entire word, a word's fragment, or only a few characters like punctuation (Song et al. 2021).

- **Lemmatization**: Given its use in numerous NLP applications, including text to speech (Zine, Meziane, and Boudchiche 2017), machine translation (Dichy and Farghaly 2003), indexing (Hammouda and Almarimi 2010), text classification (Abuhaiba and Dawoud 2017) and interactive dictionaries (Mohammed et al. 2015), lemmatization holds a significant position in the field of natural language processing (NLP).
  Each word in the text must have its related canonical word determined as part of the lemmatization process. The lemma is the simplest form of the word that conveys its primary meaning and serves as a representation of the dictionary's input.(Boudchiche and Mazroui 2019) (Khyani et al. 2021).

- **Removing stop words:** A stop word is a term that is less meaningful than other tokens. Stop words are the most prevalent terms in any natural language that have little to no or barely any semantic context in a sentence. To make other tasks easier and the primary text processing task faster, it must be eliminated as a preprocessing activity (Raulji and Saini 2016).

2.2. Topic modeling:

Topic modeling is one of the most effective text mining methods for detecting links between data and text documents and latent data. The type of data studied in topic modeling can originate from numerous sources and formats.

2.2.1. Topic modeling definition:

Finding the abstract subjects that appear in a group of documents is possible with the statistical modeling technique known as "Topic Modeling" (Alami et al. 2021).These are helpful tools made to rapidly and effectively elucidate the key themes or trends in a sizable collection of unstructured data (R. K. Gupta et al. 2022). As a result, it has been utilized more frequently to explore the primary current trends in various fields of study (Bai et al. 2021). The text documents are analyzed, and the underlying subjects (latent themes) are automatically extracted from them using the unsupervised machine learning technique known as topic modeling (Sharma, Rana, and Nunkoo 2021).

It is presumed that the words that received the highest ratings in a certain topic are those that are semantically relevant. Technically speaking, a "word" or "term" denotes the basic unit of each individual piece of data, a "document" denotes a string made up of N words, and a "corpus" denotes a collection made up of M documents that generally encompasses the whole dataset. A "topic" is represented as a probability distribution covering a specific vocabulary, and a "vocabulary" is described as the collection of all distinguishable words inside a corpus (Schöch 2021) (Vayansky and Kumar 2020).

2.2.2. Topic modeling techniques:

The most popular topic modeling techniques are:

2.2.2.1. Latent Dirichlet Allocation (LDA):

Latent Dirichlet Allocation (LDA) is a generative probabilistic model proposed by Blei et al (Blei, Ng, and Jordan 2003). It's also referred to as a three-layer Bayesian probability model that takes words, themes, and documents into account. With the use of Dirichlet distributions, it creates topic per document and words per topic models (Guo, Lu, and Wei 2021).
Figure 32 represents the graphical model of LDA. The wi,j which is the index of word w in document m, represents the input of the model. The output from the model is the K, predefined number of latent topics. Each topic k, k ∈ {1, …, K} is represented by a discrete probability distribution Φk over the vocabulary V and generated from a Dirichlet distribution Φk ~ Dirichlet(β). Additionally, every document m, m∈ {1, …, M} comes from a Dirichlet distribution Θm ~ Dirichlet(α), which is the topic distribution for each document m. From Θm we calculate zm,n per word topic assignment in document m , where β and α are the Dirichlet parameters (A. Gupta and Katarya 2021).

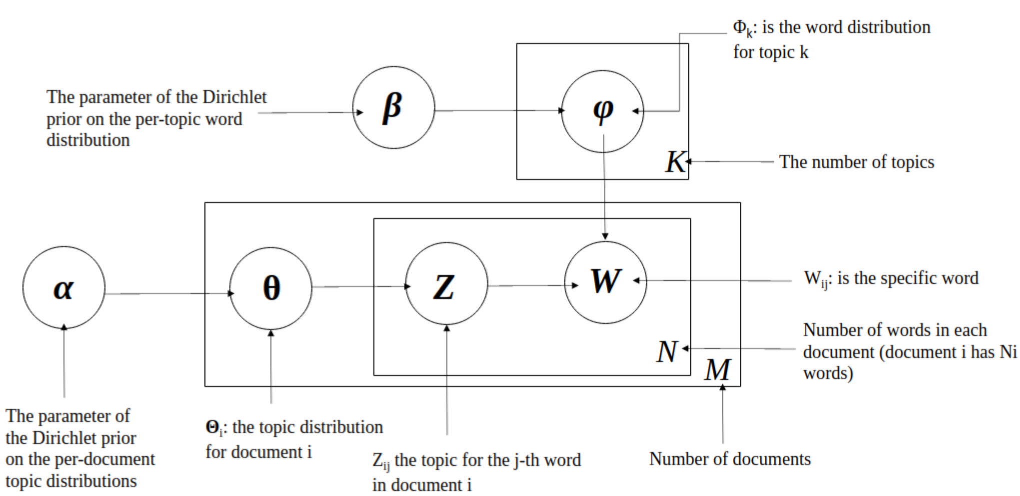

Figure 32 Generative process of LDA

2.2.2.2.Latent Semantic Analysis (LSA):

Latent semantic analysis (LSA) is a well-known technique used in computational linguistics to extract contextual and usage-based representations of words from textual corpora (Qi, Hessen, and van der Heijden 2021). In order to calculate the word count per sentence or paragraph, LSA first fills out a matrix. With each row denoting a distinct word and each column denoting a sentence or paragraph (Suleman and Korkontzelos 2021). By reducing the number of columns while maintaining the similarity structure across rows, Singular Value Decomposition (SVD), a popular dimensionality reduction technique, is applied. Finding the cosine similarity between two vectors—which goes from zero to one—allows for the matching of words (X. Yang et al. 2022).

2.2.2.3.BERTopic:

BERTopic is a topic modeling method that employs class-based TF-IDF and BERT embeddings to produce dense clusters (Grootendorst 2020). Document clustering is carried out using HDBSCAN to cluster reduced embeddings and UMAP to reduce the dimensionality of the embeddings.

Three phases are used by BERTopic to generate topic representations. Each document is first transformed using a trained language model into its embedding representation. The dimensionality of the generated embeddings is then decreased prior to clustering these embeddings in order to improve the clustering procedure. Finally, topic representations are retrieved from the document clusters using a customized class-based form of TF-IDF.

- ***Document embeddings:*** By embedding the documents, representations in vector space that may be compared semantically are produced. The embedding step is carried out by BERTopic using the Sentence-BERT (SBERT) architecture (Reimers and Gurevych 2019). Using pre-trained language models, this system enables users to transform sentences and paragraphs into dense vector representations. Instead of explicitly producing the topics, these embeddings are mostly used to group documents that share comparable semantic properties.
- ***Document clustering:*** UMAP has demonstrated that it can keep more of the local and global characteristics of high-dimensional data in smaller projected dimensions (Grootendorst 2022). As a result, document embeddings' dimensionality is decreased using UMAP. HDBSCAN is used to cluster the reduced embeddings. DBSCAN is extended to become a hierarchical clustering method in order to locate clusters with different densities (Abuzayed and Al-Khalifa 2021). HDBSCAN employs a soft-clustering technique to represent clusters, allowing noise to be modeled as outliers. This keeps unrelated papers from being grouped together and is anticipated to enhance subject representations.
- ***Topic Representation:*** Each cluster of documents will be allocated a specific theme, and the topic representations will be modeled based on those topics. Significant words for each cluster are extracted class by class using the c-TF-IDF (class-based term frequency, inverse document frequency) (Sangaraju et al. 2022). The TF-IDF allows for

a comparison of the relevance of terms across documents by calculating a word's frequency in a given document and determining the word's prevalence across the corpus (Grootendorst 2022). Conversely, if we treat every document in a cluster as a separate document and then apply TF-IDF to it, we will get relevance scores for each word in the cluster. A word cluster is more representative of a topic when its constituent words are more important.

Topic modeling algorithms can be used to identify the main themes or topics that are present in a large collection of documents, such as tweets, and represent them in a compact and meaningful way. Thus, finding and recommending tweets that are of potential interest to users from a large volume of tweets that is accumulated in real time is a crucial but challenging task.
In the following section, we will point out our contribution consisting of extracting the learners "course of interest" from the tweets they share.

3. Methodology:

Our goal is to apply NLP and Topic modelling techniques on the tweets generated by the learners in social media to identify their "real" course of interests. In the following section, we highlight in detail, the sequence of steps applied on the textual data generated by the learners to reveal their potential course of interest.

3.1. Course Topic Model architecture:

The proposed approach encompasses several steps related to the generation of topical interests as highlighted in Figure 33. The steps are performed sequentially for data preparation, data preprocessing and feature extraction. Then we train three different topic modeling methods, including: LDA, LSA and BERTopic and finally, we evaluate the performance of the models using topic evaluation metrics and validate their coherence. The sequence of steps in the proposed approach are discussed in detail below:

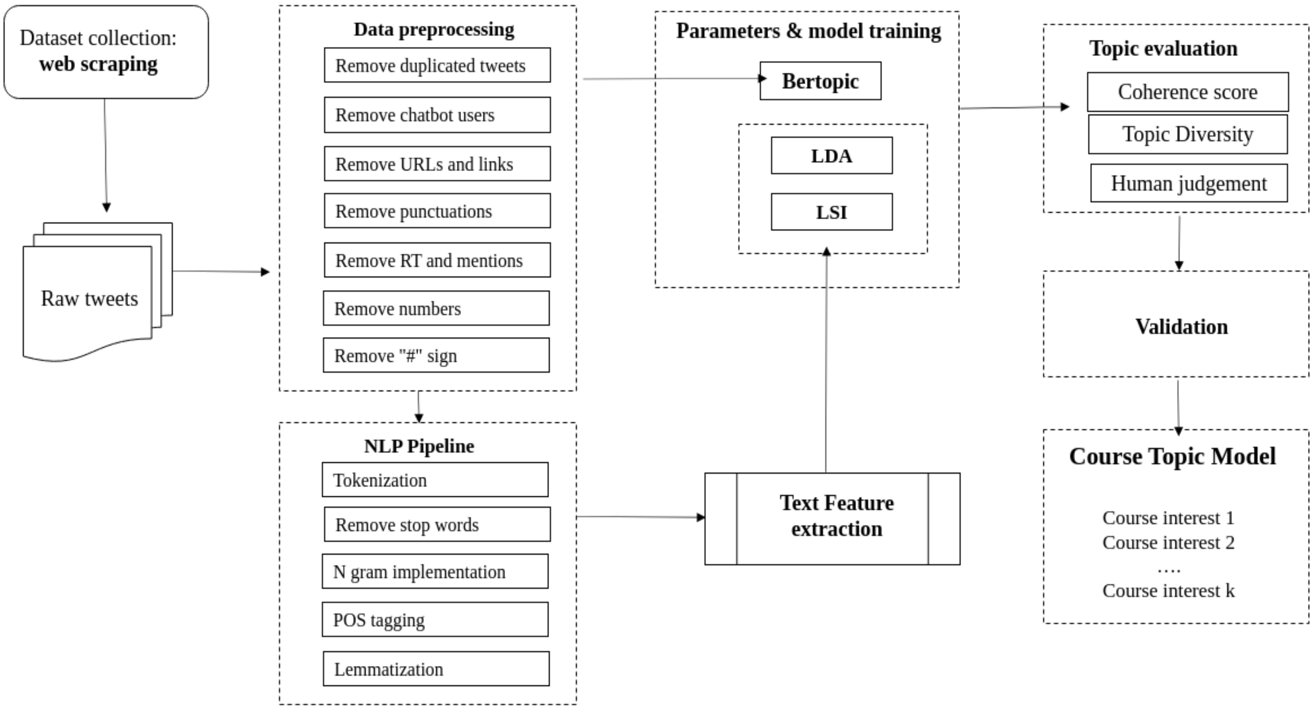

Figure 33 Sequence of steps for the proposed approach

3.2. Dataset Collection and Preprocessing:

3.2.1. Dataset collection:

A set of tweets were collected via web scraping technique for a period of two months. We used Twitter API keys and Netlytic[1] which is a cloud-based text and social networks analyzer, that allows to collect publicly accessible posts from social media.

To refine the process of scraping, we applied two filters: scarping data with English language and initiating the scarping using the keywords «Computer science» and «Artificial intelligence». A total of 2500 tweets per query were scraped. The dataset used for our experimentation contained 120 000 tweets of 12187 learners. The meta-data gathered included tweet id, learner names, tweet text, tweet type, learner id, etc. Since we are interested in the textual feature generated by the learners, we kept only learners' names and their tweets as features for our dataset.

3.2.2. Data preprocessing:

Preprocessing data is a significant step in the knowledge discovery process. Inconsistencies, missing numbers, noise, and/or redundancy are only a few of the many issues that are typically present in raw data. As a result, if subsequent learning algorithms are given low-quality data, their performance will suffer. So, by carrying out appropriate preprocessing steps, we are able to greatly affect the accuracy and dependability of subsequent automatic discoveries and decisions (Ramírez-Gallego et al. 2017).

In data preprocessing, processes including feature selection, instance selection, discretization, and other data reduction techniques are combined with data integration, cleaning,

[1] https://netlytic.org/index.php

normalization, and transformation. A final dataset that can be regarded as accurate and relevant for additional data mining algorithms is the intended outcome following a trustworthy chaining of data preparation operations (García, Luengo, and Herrera 2015).
We first cleaned our dataset from duplicated tweets and chatbot users using string pattern recognition. Each tweet should be free of links, numbers, emojis and punctuation. We used a preprocessing library called tweet-preprocessor to convert our tweets into the processed form.
Preprocessing steps included: removing URLs, removing Mentions, removing Emojis and Smileys and removing Numbers. For hashtags, we removed only the sign "#" and we kept the word after the sign, since it constitutes a valuable information about the preferences of the learner. We removed extra white spaces and punctuation as well.

3.3. NLP tasks:

The following steps were performed:

3.3.1. **Tokenization**:

We split the sentences into words while lowercasing the words, ignoring tokens that are too short (contains at least 3 letters) and removing letter accents as presented in Figure 34.

```
# Tokenization and lowercasing. we use simple_preprocess of gensim for this task
def sent_to_words(sentences):
    for sentence in sentences:
        yield(gensim.utils.simple_preprocess(str(sentence), min_len = 3, deacc=True))
#store tokens into a list
data_words = list(sent_to_words(data['text']))

data_words
```

Figure 34 Tokenization overview

We removed all non-content bearing stop words like "a", "an", "the", etc. As they do not contribute to either the representation of the tweet nor to the scoring mechanism of each tweet. We also extend the list of the stop words by high frequency words. This is done by setting a threshold in which words that occur more frequently than the threshold should be removed.
A standard stop word list by "NLTK" library has been used for this work. NLTK (Siva Rama Rao et al. 2022) is a Python module used to examine data from human language. Similar to the Stanford CoreNLP library, NLTK has an abundance of resources and offers a good number of wrappers for various programming languages.

3.3.3. **N-grams implementation:**

The majority of current NLP and its applications use N-gram-based approaches. They are frequently employed as features in vector space models, which are then represented by these features and subjected to the usual classification procedures. These features' values correspond to the n-gram frequencies, which may be weighted in some way (Sidorov et al. 2014).
We extract sequence of 'n' words that occur frequently in the corpus. In our work, we implement bi-grams and tri-grams that mean 2 words and 3 words in sequence respectively. We constructed the bi-gram and tri-gram models using The Phraser model of the Gensim library

as highlighted in Figure 35. Gensim is a Python library for topic modelling, document indexing and similarity retrieval with large corpora (Haider et al. 2020).

```
# Build the bigram and trigram models
bigram = gensim.models.Phrases(data_words, min_count=10, threshold=30) # higher threshold
trigram = gensim.models.Phrases(bigram[data_words], min_count=10, threshold=30)

bigram_mod = gensim.models.phrases.Phraser(bigram)
trigram_mod = gensim.models.phrases.Phraser(trigram)
```

Figure 35 Bi-gram and tri-gram model

Likewise known as the processing pipeline. Using part-of-speech tagging, the syntactic categories for the words in a sentence are chosen in the most probable order. It establishes the grammatical aspects of the words, including their gender, person, and other properties like part of speech and grammatical number (Tasharofi et al. 2007). The default pipeline consists of a tagger, a parser and an entity recognizer. In our work, we only use a tagger. We check part of the speech tag of each token and keep the nouns, adjectives, verbs, and adverbs using Gensim Library.

3.3.5. Lemmatization:

After removing stop words and creating bi-gram and tri-gram, we removed inflectional endings only and return the base or dictionary form of a word as it is shown in Figure 36.

```
def lemmatization(texts, allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV']):
    """https://spacy.io/api/annotation"""
    texts_out = []
    for sent in texts:
        doc = nlp(" ".join(sent))
        texts_out.append([token.lemma_ for token in doc if token.pos_ in allowed_postags])
    return texts_out
```

Figure 36 Lemmatization process

We limited our data to a sample of 10000 tweets since training the models with 120 000 tweets was time and memory consuming.

The output of this step is a clean and tokenized set of terms that will be used as an input for the following steps.

3.3.6. Text Feature extraction:

Text feature extraction is important for text categorization since it affects how accurate the results are. Text is seen as a dot in an N-dimensional space for the purposes of text feature extraction, which is based on the vector space model. One aspect of the text in digital form is represented by one of the dot's dimensions. In most cases, feature extraction algorithms employ a keyword set. The feature extraction method determines the weights of the words in the text based on these predetermined keywords, and then creates a digital vector that serves as the text's feature vector (Liang et al. 2017), (Dzisevič and Šešok 2019). In this context:

- ***A document*** refers to a single textual piece of data. This might be a book, tweet, email, text message, or song lyrics. This corresponds to a single row or observation.

- ***A corpus*** is a group of related documents. This would be the same as a complete set of rows and observations in a data set.

Some of the most popular methods of feature extraction are: Bag-of-Words and TF-IDF

3.3.6.1.Bag of Words (BoW):

A Bag of Words (BoW) is a representation model of document data, which simply counts how many times a word appears in a document (Diera et al. 2022).
BoW model assigns a vector to a document as d = (x1, x2, . . ., xl), where xi denotes the normalized number of occurrences of the i-th basis term and l is the size of the collection of basis terms. It should be noted that the basis terms are the high frequency words in a corpus, and the number of basis terms or the dimensionality of BoW vectors is less than the size of vocabulary (Zhao and Mao 2018).

3.3.6.2.Term Frequency Inverse Document Frequency (TF-IDF):

The use of the Term Frequency Inverse Document Frequency (TF-IDF) weighing scheme is an additional method for extracting features. Each word is given a weight using a TF-IDF transformation. The relative frequency of a word in a given text and the word's inverse proportion over the entire corpus, which measures a word's relevance to a certain text, are what determine the TF-IDF value (Zaware et al. 2021).

- **Term frequency (TF)** (A. Kulkarni and Shivananda 2021) :is the proportion of a word's presence in a sentence to all the other terms in that sentence. No matter how long the document is, TF captures the significance of the term. For instance, a word with a frequency of 3 in a sentence of 10 words behaves differently than a word with a frequency of 3 in a sentence of 100 words. In the first case, which is what TF does, it need to be given more weight.
  TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document).

- **Inverse document frequency (IDF):** is a log of the proportion of total rows to the number of rows in a given document containing a word: IDF = log(N/n), where N is the overall number of rows, and n is the number of rows in which the word appeared. IDF calculates a term's rarity. Words like a and the show up in all the corpus documents, but rare words are not in all documents. So, if a word appears in almost all the documents, that word is of no use since it does not help with classification or information retrieval. IDF nullifies this problem.

- **TF-IDF** is the simple product of TF and IDF that addresses both drawbacks, making predictions and information retrieval relevant. TF-IDF = TF * IDF

In our work, each tweet represents one document. We have 10 000 documents in total. We applied feature extraction using Bag of words and TF-IDF transformations to generate our corpora. In order to train both LDA and LSI, a dictionary and a corpus should be created. A dictionary encapsulates the mapping between normalized words and their integer ids. On the other hand, corpus is a list of lists containing tuples for each word id and its frequency.
While creating the dictionary, we apply the following steps:

- Filtering out high-frequency words: we used the FreqDist function to create the frequency distribution of all the words in the text. We set a threshold equal to 2000 to remove words that have frequency over it.
- Adding those words to the stop words list.
- Performing the filter_extremes[2] function that filters out tokens in the dictionary by their frequency. It has two variables: "no_below" and "no_above". Which means to keep tokens contained in at least no_below documents and no more than no_above documents. We repeated this process using respectively a set of test values:({5,0.5}, {5,0.7}, {5,1}), ({10, 0.5}, {10,0.7}, {10,1}) and ({20, 0.5}, {20,0.7}, {20,1}). The best result was obtained using no_below= 5 and no_above = 0.7.

Based on the result above, we remove all words that occur in less than 5 documents (tweets) and all words that occur in more than 70% of all the documents.
Based on the created dictionary, two corpora were generated: BOW corpus and TFIDF corpus.

3.4. Parameters and model training:

We apply LDA and LSI to both corpora. There are two types of hyperparameters —the number of latent topics for both models and Dirichlet priors (alpha and beta) for LDA.
The number of latent topics must be chosen before LDA and LSI are run. We randomly choose a value of 10 for the number of latent topics. After, we perform a grid search over this hyperparameter to find the optimal number of topics as pointed out in the next section.
Alpha and beta are the Dirichlet priors for the LDA model. We set their values to "Auto" and the model learns the right values of this parameters when it is run.
The other parameters of LDA and LSI were chosen based on a set of test values that made the models converge. Table 19 highlights the values of the different hyperparameters.

Table 19 LDA and LSA model's hyperparameters

	num_topics	chunksize	passes	iterations	alpha	eta
LDA	10	2000	20	400	auto	auto
LSI	10	2000	-	-	-	-

For BERTopic, we use a Countvectorizer transformer to eliminate English stop words and ignore terms that have a document frequency strictly higher than 60 and lower than 20 to reduce the size of the resulting sparse c-TF-IDF matrix. As a sentence-transformer model, we use

[2] https://tedboy.github.io/nlps/generated/generated/gensim.corpora.Dictionary.filter_extremes.html

BERTweet, a language model pre-trained for English Tweets. Table 20 points out the values chosen for the different parameters of BERTopic.

Table 20 BERTopic model's hyperparameters

low_memory	calculate_probabilities	verbose	n_gram_range	nr_topics
True	True	True	(1,3)	Auto

In order to prevent blowing up the memory in UMAP, we set "low_memory" to "True". The parameter "calculate_probabilities" is set to True as well to calculate the probability of a document belonging to any topic. In addition, we set the parameter "verbose" to True to track the stages of the model. For the number of topics, we set its value to "auto".

In the following section, we will describe the topic evaluation metrics used to validate the results of the trained models.

3.5. Topics evaluation and presentation:

3.5.1. Topics evaluation:

We evaluate and compare the performance between the different topic models. We use three evaluation metrics: coherence score, Inverted Rank-Biased Overlap (RBO) score and human judgement.

- ***Coherence score*** is defined as median of pairwise word similarities formed by top words of a given topic (Rosner et al. 2014). It measures the degree of semantic similarity between its high scoring words.
- ***Inversed Rank-Biased Overlap score*** evaluates how diverse the topics generated by a single model are. It compares the top N words of two topics and uses weighted ranking (Bianchi, Terragni, and Hovy 2020). A diversity close to 0 represents a redundant topic, and those close to 1 indicate more varied topics. The higher these metrics are, the better (Murakami and Chakraborty 2022).

The approach adopted to find the optimal number of topics for LDA and LSI models is to build many models (LDA and LSI) with different values of number of topics (k) and pick the one with the highest coherence value with respect to the significance of the generated topics.

The search for the optimal number of topics started with a range from two to 98, with a step of 6. During the process, only one hyperparameter varied (number of topics) and the other remained unchanged until reaching the highest coherence score.

Figure 37 presents the values of the coherence score with regard to the number of topics for LDA and LSA models with BOW corpora.

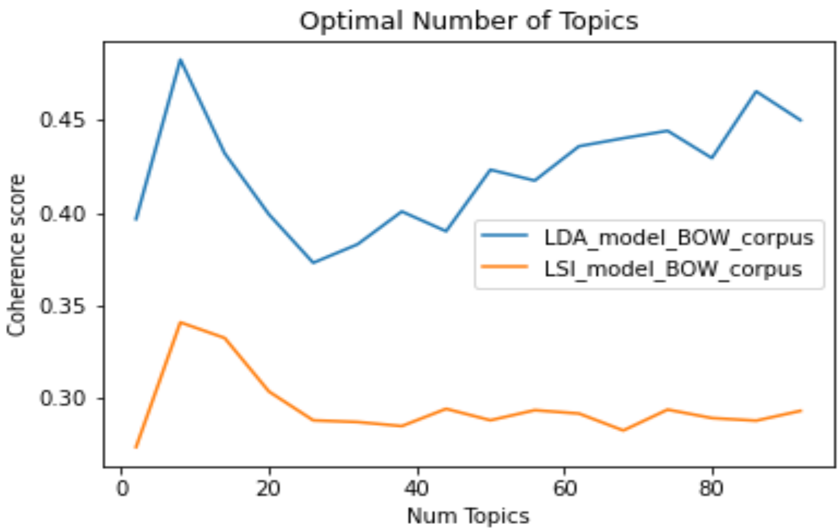

Figure 37 Optimal number of topics for LDA and LSA models using BOW corpus

Figure 38 illustrates the values of the coherence score with regard to the number of topics for LDA and LSA models with TFIDF corpora.

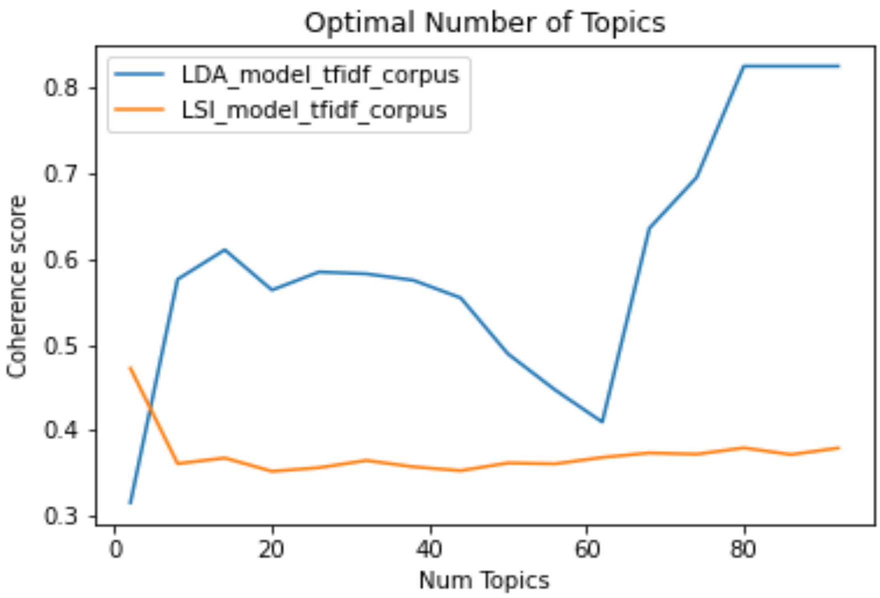

Figure 38 Optimal number of topics for LDA and LSA models using TFIDF corpus

The optimal number of topics of each model is the one that has the highest coherence score with respect to the topic's relevance and significance. The training of BERTopic model results on a number of topics equal to 40 with a coherence score equal to **0,62**. Table 21 presents the optimal number of topics for each model and the corresponding coherence score.

Table 21 Optimal number of topics with the corresponding coherence score

	BOW_LDA	**BOW_LSA**	**TFIDF_LDA**	**TFIDF_LSA**	**BERTopic**
Coherence score	0.50	0.34	0.59	0.48	0,61

Optimal number of topics	8	8	14	2	40

Based on the above table, LDA and BERTopic models have the highest coherence score compared to LSA. In order to decide which model to adopt for the generation of the course topics, we compute the RBO score of LDA and BERTopic models as highlighted in Table 22 using OCTIS, a framework for training, analyzing, and comparing Topic Models (Terragni et al. 2021).

Table 22 RBO score of LDA and BERTopic models

	LDA_BOW	**LDA_TFIDF**	**BERTopic**
RBO	1	1	0.85

3.5.2. Topics representation using Word Cloud:

To evaluate the relevance of the topics and facilitate a clear interpretation of the extracted information from a fitted LDA and BERTopic models, word cloud representation was used to generate a screenshot of the topics.

A word cloud is a kind of weighted list to visualize language or text data (Y. Jin 2017). It is employed to summarize text content by presenting a snapshot of the most commonly used words. Visual representations of text offer valuable insights, assisting users in gaining a basic understanding of the document's information without the need to read the entire text (John et al. 2018).

Figure 39 shows the visual representation of the topic of interests for the case of LDA model trained with BoW corpus.

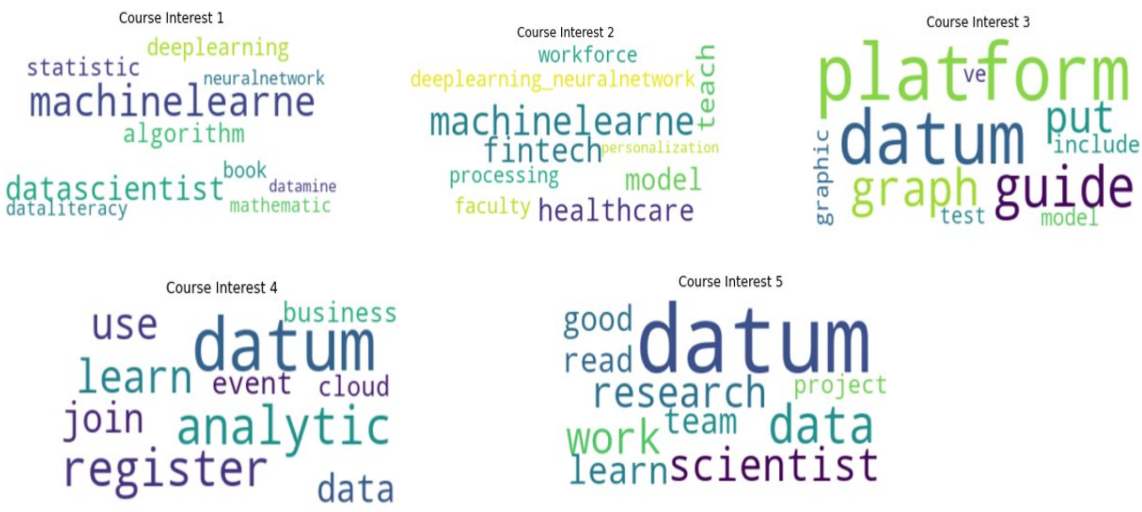

Figure 39 Word Cloud representation of LDA topics with BOW corpus

We manually labeled these topics as highlighted in Table 23:

Table 23 Labels of the course interest generated from LDA _BOW model

	Course Interest 1	Course Interest 2	Course Interest 3	Course Interest 4	Course Interest 5
LDA_BOW	Machine learning	Machine learning in fintech	Graph data platform	Data analytic	Datum research

We applied word cloud on the topics generated by the model LDA using TF-IDF corpus. Figure 40 presents the visualization of these topics:

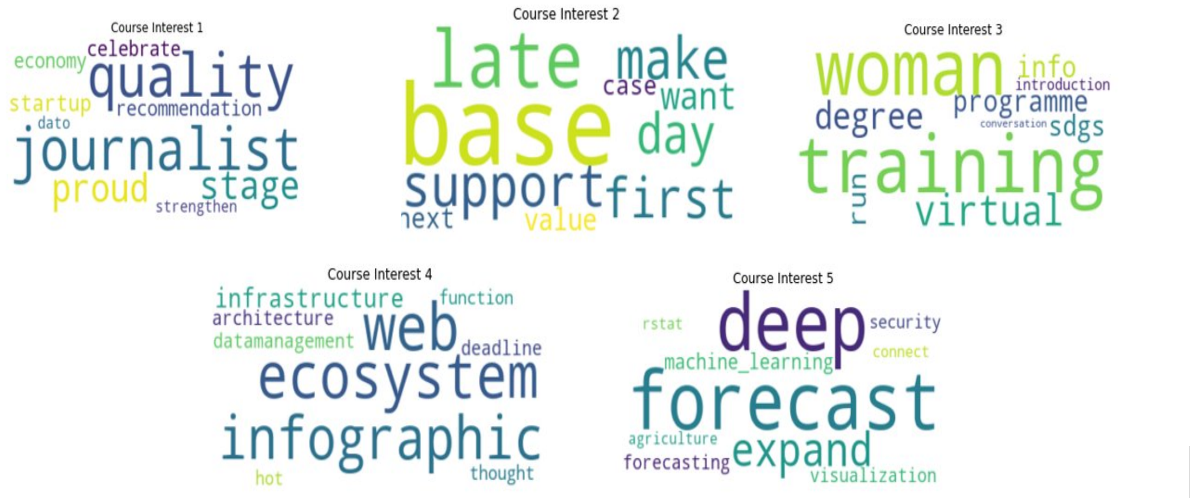

Figure 40 Word Cloud representation of LDA topics with TFIDF corpus

LDA_TFIDF	Journalism	--	Training Program for women	Digital ecosystem	Forecasting with machine learning

We visualized the topics generated by the model BERTopic as shown in Figure 41:

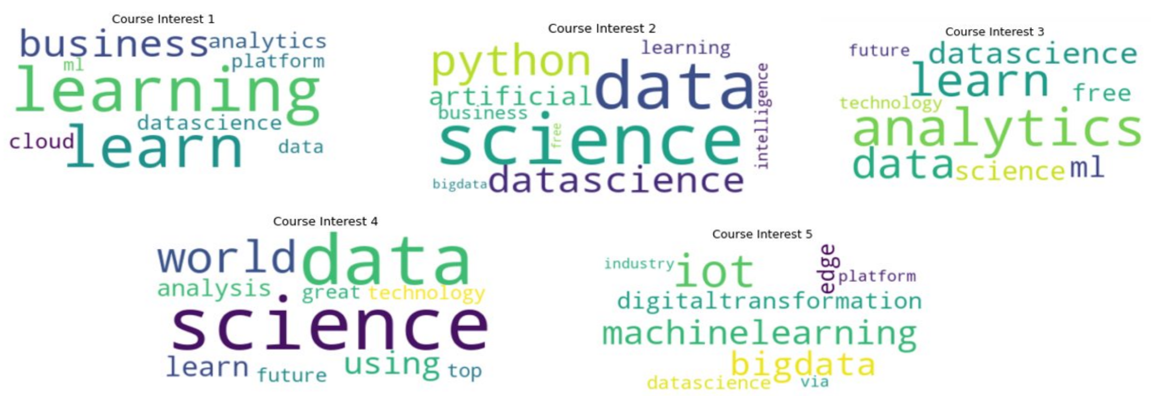

Figure 41 Word Cloud representation of BERTopic topics

Table 25 Labels of the course interest generated from BERTopic model

	Course Interest 1	**Course Interest 2**	**Course Interest 3**	**Course Interest 4**	**Course Interest 5**
BERTopic	Digital Business	Data Science with Python	Data analytics	Data science and analysis	Industrial IoT

4. Results summarization:

The aim of this work is to predict the real interest of learners from the textual content they share on their social media account to enrich their course preferences in MOOCs. Although MOOCs offer a set of courses that suit the learners predefined interest when they first sign up but those preferences do not reflect their real needs and requirements.

Our goal is to identify the course topics based on learners' interaction in Twitter, through the tweets they share. In order to do this, we require three models: LDA, LSI and BERTopic. We trained the LDA and LSI using two corpora: a bow corpus and a TFIDF corpus. We trained BERTopic using a pre-trained sentence transformer called BERTweet. We evaluated our models using the coherence score, the RBO score as well as human judgement to outline the quality and the relevance of the generated course topics.

LDA_BOW, LDA_TFIDF, and BERTopic models show prominent results with a coherence score of **0.50**, **0.59**, and **0.61** respectively, an RBO score of **1**, **1**, and **0 .86**, respectively. To decide the best model to adopt for the course topics, we judge the relevance and the significance of the generated topics as well. We manually labeled the first five topical interests as highlighted in tables 20,21 and 22.

Based on the tables discussed in section 'Topics evaluation ', BERTopic outperformed LDA regarding the quality and the relevance of the topical interest that are more representative for the course of interest of learners.

For each tweet, we identify the dominant topic that has the highest probability, and then for each learner, we get the top two dominant topics. Table 26 points out the top two courses of interest for four learners.

Table 26 Top 2 course of interest of four learners

Learner ID	**Top 2 courses of interest**
ID 1	Digital business and data analytics
ID 2	Data science and machine learning
ID 3	Industrial IoT and digital business
ID 4	Artificial intelligence and data science

5. Conclusion:

In conclusion, this chapter presents a comprehensive approach for identifying learners' course interests based on their interactions on social media, particularly Twitter. The proposed approach involves several steps, including dataset collection and preprocessing, NLP pipeline, text feature extraction using Bag of Words (BoW) and TF-IDF transformations, and training

three topic modeling models: Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and BERTopic.
We first collected a large dataset of tweets from learners, cleaned and preprocessed the data to remove duplicates, links, emojis, and other noise. The NLP pipeline was applied to tokenize and remove stop words from the tweets, followed by implementing N-grams and part-of-speech tagging. Lemmatization was also performed to convert words to their base form.
After preprocessing, text feature extraction was carried out using BoW and TF-IDF transformations to create corpora. The LDA and LSA models were trained on these corpora with varying numbers of topics to find the optimal value for coherence. Similarly, BERTopic was trained using the CountVectorizer transformer and BERTweet as a sentence-transformer model.
We then evaluated the performance of the models using coherence scores, Inverted Rank-Biased Overlap (RBO) scores, and human judgement. The results showed that BERTopic outperformed LDA and LSA in terms of coherence and relevance of generated topics.

Chapter 5: Application of Recommender systems in e-learning

1. Introduction:

The rapid expansion of the computer industry, particularly the Internet, has fundamentally altered our way of life. Users can engage in a variety of activities with just a few clicks, including news browsing, online shopping, e-learning, entertainment, and etc. (Fayyaz et al. 2020). Recommendation Systems (RS) have arisen to help users find pertinent content and services on the web. However, users frequently feel overwhelmed by the abundance of options (Jannach et Jugovac 2019).

In the context of e-learning, recommendation systems play a pivotal role in assisting learners to discover relevant and valuable learning materials that satisfy their specific needs. These systems are often employed to address the issue of low engagement, which arises from the generic "one-size-fits-all" approach that disregards the significance of individual distinctions in teaching and learning processes. However, the challenge lies in providing learning resources personalized to each learner, given the varied characteristics they possess, including their learning preferences, existing knowledge, and the way they progress in their learning journey.

Current recommendation methods often fail to consider these individual learner attributes. To solve this problem, a potential solution involves incorporating more personalized information about each learner into the recommendation process. This refinement promises to ensure the delivery of customized learning materials that meet the unique learning needs of individuals, even when dealing with diverse learning profiles(Klašnja-Milićević et al. 2011) (Tarus, Niu, and Mustafa 2018a).

2. Types of recommender systems:

There are several different definitions that have been presented for recommender systems. A general definition of these systems is provided by Robin Burke, who describes them as "Systems that can deliver personalized suggestions to direct users to interesting and valuable information in a wide data space." (BURKE 2002). As well as this, (Lu et al. 2015) described recommendations system as software and methods for making suggestions to users. The term "item" in this sense refers to what the system suggests, which can include products to buy, films to watch, training materials, services, and much more, offering a vast range of options. With the Tapestry system, they made their debut in 1992 (Goldberg et al. 1992).

There are different formulations for the recommendation problem, with two primary models being (Aggarwal 2016):

- Prediction version of the problem: The goal is to estimate the rating value associated with a user-item pair. This is accomplished by leveraging historical training data that holds insights into user preferences for different items.
- Conversely, in **the ranking version of the problem**: The focus shifts from predicting exact user ratings for specific items. Instead, it aims to recommend the top-k items for a particular user or identify the top-k users interested in a specific item. This approach

proves more practical for businesses and retailers seeking to offer personalized recommendations without the necessity of precise rating predictions.

Historically, the three types into which existing approaches for recommendation systems can be broadly divided are: ***Content Based (CB) Methods, Collaborative Filtering (CF) Methods and Hybrid Methods*** as highlighted in figure 43.

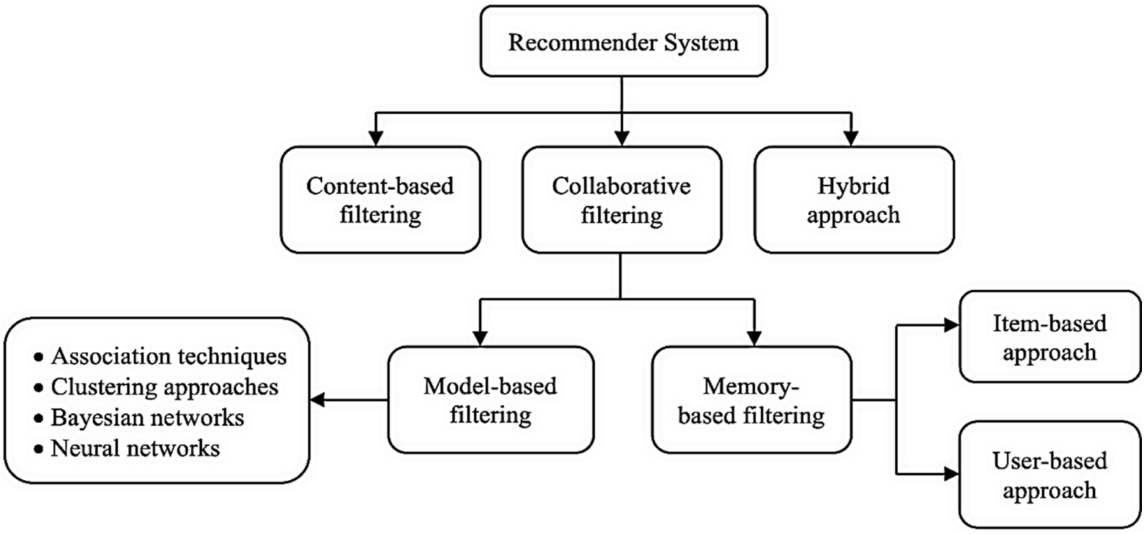

Figure 43 Types of the most used Recommender systems

2.1. Content Based Methods:

Content-based (CB) (Guruge, Kadel, and Halder 2021)**:** The recommendation engine learns to suggest items with features that are comparable to those the user previously liked in terms of content.

The CB recommender system makes suggestions for products based on the item's description that would interest the user or be similar to their previous favorites. Typically, this recommendation process involves comparing item characteristics with the user's preferences. These user preferences can be inferred explicitly or implicitly through surveys or by tracking the user's transactional behavior over time. The similarity between items is established based on the attributes of the items being compared as presented in figure 44.

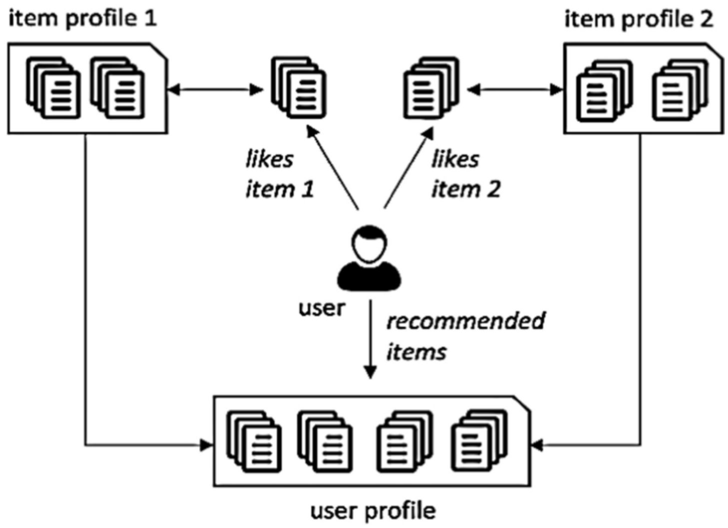

Figure 44 Content Base recommender System

Content based recommenders can be classified into:

- **Case based reasoning techniqu**e that suggest products that are most likely to be enjoyed by the learner before.
- **Attribute-based technique** that suggest products based on how well their characteristics match the learner profile.

2.2. Collaborative Filtering (CF) based methods:

Recommends items based on past preferences of other users with similar tastes to the active user. In order to measure taste similarity between two users, the rating history is compared, and the rating measures the level of interest a user has in an item (Alhijawi and Kilani 2020) (Ghenname 2015). The k Nearest Neighbors (kNN) algorithm is the most popular collaborative filtering algorithm: Nearest neighbors are users with similar tastes with the active user in terms of ratings similarity.
Collaborative recommendation algorithms can be classified into two types, commonly referred to as memory-based and model-based algorithms (Koren, Rendle, and Bell 2021) (Adomavicius and Tuzhilin 2005):

- **Memory-based:** known also as neighborhood-based techniques (Sarwar et al. 2001). This method allows for the discovery of target user neighbors who have interests with the system's content. This closeness of interests is based on system item assessments by users. As a result, once neighbors are located, notes and associations can be merged in a variety of ways to provide recommendations to the target user (Mittal 2016). Within memory-based collaborative methods, there are two categories: ***user-based collaborative filtering*** and ***item-based collaborative filtering***.

In the user-based approach, to calculate a user's rating for a new item, the system identifies other users in the user neighborhood who have rated that same item previously. If the new item receives favorable ratings from this user neighborhood, it is subsequently recommended to the user as pointed out in figure 45.

In the item-based approach, a set of similar items that the user has previously rated is constructed, known as the item-neighborhood. To predict the user's rating for a different new item, this approach calculates the weighted average of all the ratings within the similar item-neighborhood as shown in figure 46.

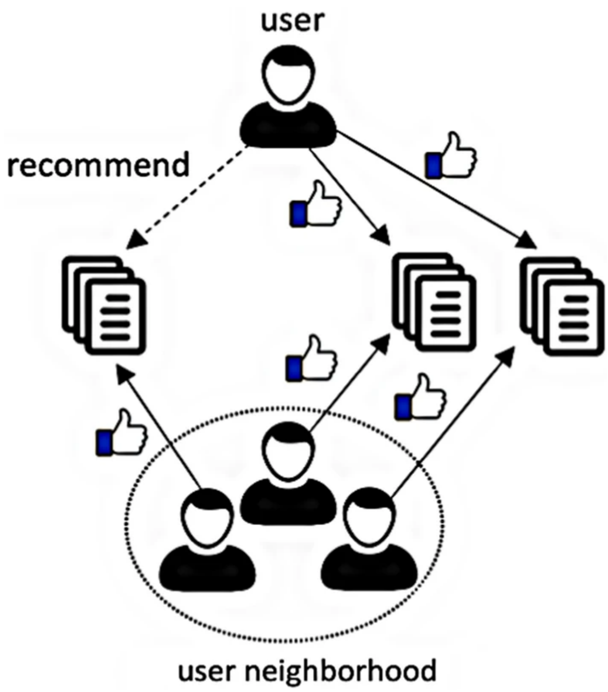

Figure 45 User-based approach

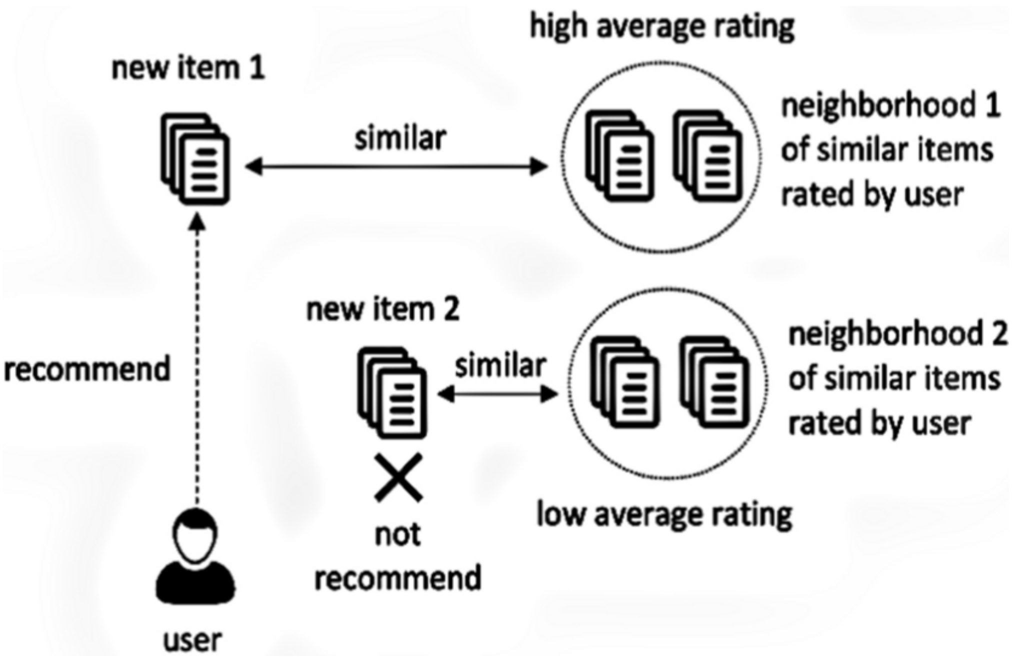

Figure 46 Item-based collaborative filtering

- **Model-based:** in this approach the correlation is measured between the content instead of measuring it between the users. The basic idea is to look for items similar to the items that a user has already chosen in the system in order to recommend them. Preliminary processing is performed on an evaluation matrix in order to select similar items and thus make recommendations in real time contrary to the user-centric approach which is very greedy in terms of memory resources (Ghenname 2015).

2.3. Hybrid Method:

The hybrid recommendation system – as the name implies –integrates two or more recommendation algorithms in various ways to capitalize on their complimentary benefits (Mittal 2016) and (Beel et al. 2016). A hybrid algorithm can combine outcomes from different methods, or it can employ content-based filtering in a collaborative approach, or use a collaborative filtering method within a content-based approach. This fusion of diverse techniques typically leads to enhanced performance and increased accuracy in numerous recommender system applications.

Table 27 presents the different recommender system techniques with their representative algorithms along with their advantages and disadvantages.

Table 27 Types of recommender systems

Techniques	***Representative algorithm***	***Advantages***	***Disadvantages***
Memory Based Collaborative Filtering	1.User Based CF 2. Item Based CF	1. Simple to implement 2. It's easy to add data. 3. not to worry about content 4. Reliable scaling	1. Relies on direct feedback 2. The cold start issue 3. The sparsity issue 4.Non scalable for large datasets
Model Based Collaborative Filtering	1. Slope-one CF 2. Matrix factorization	1. Increase prediction accuracy 2.Enhance scalability and the sparsity issue	1. The model is pricey 2. Information loss during matrix factorization
Hybrid Collaborative Filtering	Combination of memory based and model based	1. Get over sparsity's constraints 2. Increase prediction accuracy	1. Growing complexity 2. Implementation challenges
Content Based Filtering	Content Based filtering algorithm	1. There are no limitations or cold starts.	1.Items must have rich descriptions.

	using Hidden Markov Model	2. It guarantees privacy.	2. Demands a structured user profile 3. Overspecialization in content is a problem
Hybrid Filtering	Combination of Collaborative and Content Based algorithm	1. The advantages and disadvantages of content-based and collaborative methods are complementary. 2. Address the issue of shortage and cold starts	Difficult to implement

The most often, recommendation systems track user activity, such as rating, clicking, purchasing, and commenting. User profiles or product descriptions are used in content-based strategies to produce recommendations. Without leveraging user or product content information, collaborative filtering-based systems rely on previous behaviors or preferences, such as user ratings on products. By integrating methods based on collaborative filtering and content-based, hybrid approaches aim to achieve the best of both worlds. (P. V. Kulkarni, Rai, and Kale 2020)

There have been a variety of existing Recommender Systems and recommendation techniques, among them:

- **Knowledge-based (KB):** These systems give users recommendations for products based on their understanding of how those products fit their needs. They use three different kinds of information: knowledge of the users, knowledge of the items, and knowledge of how well the item satisfies the users ’requirements. Knowledge-based techniques are applied in the process of making recommendations in the context of e-learning by combining knowledge about the learners and the course materials (Khanal et al. 2020).

- **Demographic-based (DB) systems:** seeks to classify users according to personal characteristics and provide suggestions based on demographic groups. Individuals that share or have comparable personal characteristics are likely to have similar item preferences (Fayyaz et al. 2020), (BURKE 2002).

- **Context aware-based (CA):** All information that can be utilized to describe an entity's situation is considered context aware (CA). Before using a conventional recommendation algorithm, the data set is filtered using contextual information in a context-aware method (D. Nawara and R. Kashef 2021) (J. Kim, Lee, and Chung 2014)

- **Trust-aware based (TA):** aims to produce individualized suggestions using well-known perspectives and trusted relationships. Trust plays a key role in e-learning and e-commerce recommender systems because it strengthens user relationships in social networks (Deebak and Al-Turjman 2020)

- **Ontology-based (OB) recommenders:** Are knowledge-based recommender systems that employ ontologies for knowledge representation. Ontologies are used to infer user interests as well as enhance user profile in the field of recommender systems (Tarus, Niu, and Yousif 2017a)

3. Application of recommender systems in e-learning:

3.1. E-learning recommender system:

It can be difficult to suggest learning resources to specific learners, in the context of e-learning due to the variety of characteristics they have, including background knowledge, history, competency level, learning style, and etc. Even if two learners receive similar ratings, their individual characteristics require for different recommendations. This emphasizes how crucial it is for recommender systems in e-learning to take into account each learner's specific preferences and needs. As a result, personalization becomes essential to generating practical recommendations in e-learning contexts. For each unique learner, it is crucial to carefully analyze learners' characteristics in order to ensure successful recommendations.

In the literature, there are several works covering recommender systems in e-learning.In (Jena et al. 2023), authors developed a recommender system for e-learning course recommendations using collaborative filtering, incorporating machine learning models like K-nearest neighbor (KNN), Singular Value Decomposition (SVD), and neural network-based collaborative filtering (NCF) models. The analysis utilizes a dataset containing one lakh Coursera course reviews from Kaggle. The system aids learners in selecting e-learning courses based on their preferences. The implementation is done in Python, and the models' performance is evaluated using metrics like hit rate (HR), average reciprocal hit ranking (ARHR), and mean absolute error (MAE). Similarly, the work presented in (Talaghzi et al. 2023), presented a course recommender system for the E-Dirassa platform, designed to serve both new and active learners. The system utilizes three distinct recommendation approaches: content-based recommendation and course-based collaborative filtering for active learners, and static profile-based collaborative filtering for new learners. It utilizes multi-criteria decision aid (MDA) techniques to choose the appropriate textual similarity measure for the content-based recommendation approach. In another work (T. Liu et al. 2022), authors conducted an in-depth examination of existing recommendation systems in e-learning environment based on the collected literature, leading to the presentation of an overall course recommendation system framework. The paper presented a systematic review of deep learning-based recommendation systems. It begins by introducing recommendation systems in e-learning and providing a comprehensive survey and classification of deep learning techniques used for course recommendation. A complete literature review of ontology-based recommenders for e-learning was conducted as well as part of the work mentioned in (Tarus, Niu, and Mustafa 2018b). In order to give an overview of the developments in ontology-based recommendation for e-learning, authors first evaluated and categorized journal papers published between 2005 and 2014. Second, they organized and highlighted the broad range of ontology-based e-learning recommender systems by classifying them. Thirdly, they investigated the approaches of knowledge representation, various types of ontologies, the ontology representation languages, and the learning resources that these systems use. Finally, they discussed the future trends of this recommendation approach in the context of e-learning.

Furthermore, (Amane, Aissaoui, and Berrada 2022) introduced an e-learning recommender system adopting a dynamic ontology approach. The suggested approach makes use of semantic descriptions of courses and students, combining clustering algorithms with collaborative and content-based filtering strategies to get the top N recommendations. Experiments are conducted implementing a combination of the well-known "Coursera" dataset and the university's USMBA dataset to assess the performance of the system. The outcomes showed that the suggested method was more effective than content-based approaches in the recommendation process. Moreover, an ontology-based (OB) content recommender system was also proposed by authors in (Jeevamol and Renumol 2021) in order to address the issue of the new user cold-start problem. Ontology is used to effectively model learners and learning items in the suggested recommendation model, capturing their distinctive qualities. The recommendation model incorporates collaborative and content-based filtering algorithms to produce the top N recommendations based on learner evaluations. Similarly, a semantic framework based on ontology was established in the work given in (Joy, Raj, and V. G. 2021) to efficiently address the pure cold-start problem in content recommenders. Both domain knowledge about learners and Learning Objects (LOs) are included in the ontology. A semantic model is built by combining different critical learner factors, such as learning style, knowledge level, and background knowledge.

Authors in (Khanal et al. 2020) thoroughly evaluated the current recommendation approaches used in e-learning including Content-Based, Collaborative Filtering, Knowledge-Based, and Hybrid Systems. The paper provides a taxonomy of recommender systems, along with machine learning algorithms, evaluation metrics, and insights on challenges and issues that require more research. This study revealed that Collaborative Filtering was a widely utilized recommendation method in online learning, with most studies seeking to raise the standard of suggestions.

In the study conducted by (Dahdouh et al. 2019), authors introduced a distributed recommendation system specially designed for e-learning platforms, using advanced big data technologies. They made use of a powerful parallel FP-growth algorithm available within the Spark Framework and the Hadoop ecosystem. This system works by finding connections between what students do, using a method called association rules. It also looked at historical data about course enrollments and students' activities to find interesting patterns in the records. These patterns were then used to create a list of courses that would suit students better, based on their behaviors and what they like.

In (Tarus, Niu, and Kalui 2018), a mixed recommendation method was presented to give personalized learning resource suggestions to learners. This approach combines three key elements: context awareness, sequential pattern mining (SPM), and collaborative filtering (CF) algorithms. Context awareness is used to include important information about each student, like what they know and what they aim to learn. The SPM algorithm digs through web logs to find out how students access information in a particular order. Then, by using CF, the system creates predictions and suggestions based on this contextual information and the sequence in which students access content.

Authors in (Benhamdi, Babouri, and Chiky 2017) introduced a new recommendation system, called NPR_eL (New multi-Personalized Recommender for e Learning), which combines two techniques: collaborative filtering and content-based filtering. This method was incorporated into a learning environment to provide students with learning materials that are personalized to their needs. This approach provides learners with the best learning materials based on their preferences, interests, background knowledge, and memory capacity.

Throughout the existing literature, numerous studies have tackled the challenges of e-learning recommendation systems, employing a variety of approaches to enhance the learning experience. These studies have explored collaborative filtering, deep learning techniques, ontology-based recommendation systems, and the integration of context awareness, sequential pattern mining, and more. Some have investigated specific issues such as the new user cold-start problem, while others have conducted comprehensive evaluations of existing recommendation methods.

Recommending appropriate learning resources to individual learners within e-learning poses a challenging task due to the wide array of characteristics that each learner possesses, including their background knowledge, educational history, skill levels, learning preferences, and more. Even if two learners seem similar based on certain criteria, their unique attributes often necessitate different recommendations. This underscores the critical importance of personalization in e-learning recommender systems, which must consider each learner's specific preferences and requirements to provide effective suggestions. Therefore, a deep understanding of each learner's characteristics is crucial to ensure successful recommendations.

3.2. Recommender System based on learner profile:

E-learning recommender systems, using learner profiles, are vital for personalized learning. They adapt educational content to each learner's unique needs, enhancing engagement and outcomes. These systems create more effective, learner-centric online education. Many works have highlighted the development of e-learning recommender systems using learner profile information.

The work presented in (Amane, Aissaoui, and Berrada 2022), developed an e-learning recommender system utilizing a dynamic ontology approach. The proposed system employs semantic descriptions of courses and learners, integrating collaborative and content-based filtering techniques with clustering methods to generate top N recommendations. To evaluate the systems' performance, experiments are conducted using a combination of the well-known Coursera dataset and the university's USMBA dataset. The results demonstrated the effectiveness of the proposed method in the recommendation process, outperforming content-based methods. In another work (Imran et al. 2016), authors in, introduced a personalized learning material recommendation system. This system aims to help learners by suggesting particular learning materials within a course. It achieves this by examining the material a learner is presently engaged with and also taking into account materials that have proven useful to other learners with similar interests. This personalization has the potential to enhance the overall learning experience by directing learners to valuable materials they might not have thought of initially.

Further, the work presented in (Dwivedi and Bharadwaj 2015) , focused on the challenge of recommending educational resources for a group of learners rather than for individual learners. The fundamental issue in group recommendation concerns the individual preferences held by different learners within a group, leading to the development of a collective learner profile, named the Unified Learner Profile (ULP), which accurately reflects the combined preferences of all learners. Initially, the authors presented a technique for integrating individual profiles to produce the ULP taking into account learning preferences, educational backgrounds, and individual learner evaluations. In order to improve group recommendations, the authors then presented a collaborative approach combined with ULP. The study offers experimental results that demonstrate the viability of the suggested group suggestion approach for e-learning.

Furthermore, authors in (Bourkoukou and Bachari 2018), introduced a new adaptive learning system called LearnFitII, which can automatically adjust to the changing preferences of learners. This system identifies various learning styles and learner habits by assessing the psychological model of learners and examining their server logs. First, it creates a personalized learning plan to address the challenge of starting with learners who have no prior data by using the Felder and Silverman model. Then, it analyzes learners' habits and preferences by mining information about their actions and interactions. Finally, the learning plan is reviewed and updated using a hybrid recommendation system that combines K-Nearest Neighbors and association rule mining algorithms. The results of testing this system in real educational settings demonstrate that taking into account the learner's preferences enhances the quality of learning and satisfies the learner. In another work (Tarus, Niu, and Yousif 2017b) , authors presented a hybrid knowledge-based recommender system that combines ontology and sequential pattern mining (SPM) to recommend e-learning resources to learners. The recommendation approach relies on ontology to model and represent domain knowledge related to learners and learning resources, while the SPM algorithm identifies sequential learning patterns of learners. The method involves four key steps: (1) the creation of an ontology to capture knowledge about learners and learning resources, (2) the computation of ratings similarity based on ontology domain knowledge to predict preferences for the target learner, (3) the generation of the top N recommended learning items through collaborative filtering, and (4) the application of the SPM algorithm to the top N learning items to produce final recommendations for the target learner. Authors conducted several experiments to assess the performance of the hybrid recommender system, and the results indicate improved effectiveness. Furthermore, the hybrid approach offers solutions to both the cold-start problem and data sparsity issues by leveraging ontological domain knowledge and the sequential access patterns of learners, even before initial data becomes available in the recommender system.

Moreover, in this research paper (Tarus, Niu, and Kalui 2018), authors developed a hybrid recommendation approach that merges context awareness, sequential pattern mining (SPM), and collaborative filtering (CF) algorithms to suggest learning resources to learners. The recommendation strategy leverages context awareness to consider learner-specific details like their knowledge level and learning objectives. The SPM algorithm delves into web logs to uncover the sequential patterns in a learner's access to resources. CF then utilizes this contextualized information and the learner's sequential access patterns to make predictions and generate recommendations. The evaluation of this hybrid approach reveals that it can surpass other recommendation methods in terms of the quality and accuracy of the recommendations provided.

Additionally, The work highlighted in , developed a MoocRec.com which is an online platform designed to help learners discover courses that align with the skills required for their dream job positions. MoocRec's recommendation system leverages a combination of Matrix Factorization (MF) and Collaborative Filtering (CF) algorithms. These algorithms make use of external data sources, including user skills and course attributes, to forecast course trends and provide predictions for course ratings based on these trends. Also, authors in (Pireva and Kefalas 2018), introduced a Recommender System based for Cloud e-Learning (known as CeLRS). CeLRS employs hierarchical clustering to identify the most suitable resources and employs a vector space model to arrange these resources in terms of their significance for each individual learner.

To summarize, traditional e-learning platforms such as EdX, Coursera, Udemy and etc., employ methods like collaborative filtering, content-based or hybrid approaches to support recommendation systems. However, despite all the efforts made to suggest suitable courses, the former platforms may not recognize the appropriate needs of the different learners. Thus, it is essential to suggest recommendations that consider the characteristics of individual learner profiles.

In response, our thesis holds significant promise for enhancing personalized recommendations in the area of e-learning, aligning educational content more closely with individual learner needs. Our approach empowers recommendation systems by taking learner profiles into account and integrating information about the learner from his social media profile. We achieve this by combining both the outcomes of the social profile ontology and the course topic model. Based on our results, The MOOCs platforms could construct a hybrid recommender system using both knowledge and content base methods to identify and suggest relevant educational resources tailored to the learner's profile.

Conclusion and Perspectives

E-learning has transformed education since it offers convenient and adaptable learning alternatives especially during the COVID 19 pandemic and it has proven to be durable and cost-effective even after the quarantines. As a result, e-learning has become a crucial component of modern education, providing accessible, inclusive, and interesting learning opportunities for learners all over the world.

Additionally, Massive Open Online Courses (MOOCs) as e-learning platforms, provide free access to educational materials, improving opportunities for independent learning and personal development. However, MOOC dropout rates is still persisting due to reasons such as lack of continuous engagement, impersonalized content recommendation and less social interactivity.

Further, the dominance of social media platforms (e.g., LinkedIn, Twitter, YouTube, Facebook, Instagram, TikTok, etc.) keeps growing. This is due to its connectivity, interactivity, and user-generated content which facilitates the sharing of information by reflecting users' sentiments, thoughts, and actions.

After reviewing the factors behind drop-out in MOOCs using literature, and highlighting the potential of using social media features to promote learner engagement, we aim to exploit the learner ‘s profile social information to enrich his profile within MOOC for a personalized learning experience. Furthermore, Personalizing learner experience in MOOCs is challenging due to the diverse characteristics the learners possess, such as interest, courses preferences, competency level, learning style, and etc.

This thesis explores the intersection of MOOCs and social media using AI-driven approaches. Our research applies ontologies and machine learning to improve the e-learning experience in order to meet the needs and preferences of individual learners’ profiles. Specifically, we have implemented a social learner profile ontology to create a more comprehensive view of the learner.

We aim to integrate information from different sources (MOOC and social media) and combine them to gain a more complete understanding of each learner's interests, abilities, and needs.

For this purpose, we created two local ontologies to represent the learner profile in MOOC and the user profile in social media. Based on the well-known standards and ontologies such as IMS LIP, FOAF, Organization ontology, SIOC ontology and etc., we proposed our proper attributes to create a more customized and personalized profile. Then, we rely on ontology engineering techniques such as ontology mapping and merging to implement our social profile ontology.

Using the previous ontology, we focused on the ‘Interest’ component, to detect learners' potential topics of interest from their content on social media (tweets). We implemented NLP pipeline and applied topic modeling algorithms such as LDA, LSA and BERTopic on the Tweets. This could provide information on educational resources that fit better the learner profile.

Our thesis is a work in progress that has given some promising results, including our commitment to sharing the ontology with the community, ensuring its 'reusability' and providing comprehensive documentation for accessibility.

List of publications:

Parts of this thesis have been published in the following referenced journals and conferences:

- Zankadi, H., Hilal, I., Daoudi, N., & Idrissi, A. (2018, May). Facebook and MOOCs: a comparative analysis for a collaborative learning. In 2018 6th International Conference on Multimedia Computing and Systems (ICMCS) (pp. 1-7). IEEE. Rabat, Morocco.

- Zankadi, H., Hilal, I., Daoudi, N., & Idrissi, A. (2018, December). Building an ontology-based user profile for a social learning environment. In 2018 1st International Conference on Modern Intelligent Systems Concepts (MISC). Rabat, Morocco

- Zankadi, H., Hilal, I., & Idrissi, A. (2019, October). Designing a Middleware course for a real time interactive learning in Social Learning Environment. In 2019 1st International Conference on Smart Systems and Data Science (ICSSD) (pp. 1-5). IEEE. Rabat, Morocco.

- Zankadi, H., Hilal, I., Daoudi, N., & Idrissi, A. (2019, December). Towards a social learning environment. In Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS) (pp. 607-610). Munich, Germany.

- Zankadi, H., Hilal, I., Idrissi, A., & Daoudi, N. (2022). A Social Profile Ontology to Enhance Learner Experience in MOOCs. International Journal of Emerging Technologies in Learning, 17(4).

- Zankadi, H., Idrissi, A., Daoudi, N., & Hilal, I. (2022). Identifying learners’ topical interests from social media content to enrich their course preferences in MOOCs using topic modeling and NLP techniques. Education and Information Technologies, 1-18. Springer.

- Er-Rafyg, A., Zankadi, H., & Idrissi, A.,AI in Adaptive Learning: Challenges and Opportunities, (2024), Accepted and presented in International Conference on Modern Artificial Intelligence and Data Science Systems (MAIDSS 2024)

Perspectives:

Accordingly, it is interesting to pursue some perspectives as follows:

- Using a real MOOC platform, such as EDX or Coursera, to enhance their recommendation system with the output of our approach, in order to quantify its impact on the dropout rate and the learners' engagement and course completion.

- Ensuring inclusive education and lifelong learning opportunities taking into account a more personalization of the profile for learners with special needs such as persons with a hearing or visual impairment and persons with Attention-Deficit/Hyperactivity Disorder (ADHD) and autism spectrum disorder (ASD) disorders.

- Exploring the differentiations in behavior vis-à-vis the learning practices, regarding the different generations such as Baby Boomers, Generation X, Millennials, Generation Z, and Gen Alpha. Discovering these particular differentiations could enrich and extend the personalized profile proposed by our approach to better meet their expectations.

- Additionally, the current revolution powered by large language models like GPT-3.5 could enhance learner support and fostering a more interactive and dynamic learning experience. This might be enriched by integrating ChatGPT into MOOC platforms, to engage a personalized conversations based on the social profile provided by our approach. Therefore, our approach will empower its natural language processing capabilities, which enables more personalized and contextually relevant responses for MOOCs' learners.

- Finally, the potential of incorporating Virtual Reality (VR) combined with our approach could be investigated to enhance the learner's engagement and knowledge retention.

References:

Abdillah, Leon Andretti. 2017. « Enriching Information Technology Course Materials by Using Youtube ». *arXiv preprint arXiv:1708.04878*.

Abel, Fabian, Qi Gao, Geert-Jan Houben, et Ke Tao. 2011. « Semantic enrichment of twitter posts for user profile construction on the social web ». In , 375-89. Springer.

Abuhaiba, Ibrahim SI, et Hassan M Dawoud. 2017. « Combining different approaches to improve arabic text documents classification ». *International Journal of Intelligent Systems and Applications* 9 (4): 39.

Abuhammad, Sawsan. 2020. « Barriers to Distance Learning during the COVID-19 Outbreak: A Qualitative Review from Parents' Perspective ». *Heliyon* 6 (11): e05482. https://doi.org/10.1016/j.heliyon.2020.e05482.

Abuzayed, Abeer, et Hend Al-Khalifa. 2021. « BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique ». *Procedia Computer Science* 189: 191-94.

Adomavicius, G., et A. Tuzhilin. 2005. « Toward the next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions ». *IEEE Transactions on Knowledge and Data Engineering* 17 (6): 734-49. https://doi.org/10.1109/TKDE.2005.99.

Aggarwal, Charu C. 2016. *Recommender systems*. Vol. 1. Springer.

Ahmadaliev, Doniyorbek, Chen Xiaohui, Murodjon Abduvohidov, Asilbek Medatov, et Gulbahor Temirova. 2019. « An adaptive activity sequencing instrument to enhance e-learning: an integrated application of overlay user model and mathematical programming on the Web ». In *2019 International Conference on Computer and Information Sciences (ICCIS)*, 1-4. https://doi.org/10.1109/ICCISci.2019.8716473.

Ahn, Janet N., Danfei Hu, et Melissa Vega. 2020. « "Do as I Do, Not as I Say": Using Social Learning Theory to Unpack the Impact of Role Models on Students' Outcomes in Education ». *Social and Personality Psychology Compass* 14 (2): e12517. https://doi.org/10.1111/spc3.12517.

Akbari, Mohammad, Xia Hu, Nie Liqiang, et Tat-Seng Chua. 2016. « From tweets to wellness: Wellness event detection from twitter streams ». In . Vol. 30.

Al Mulhem, Ahmed. 2020. « Investigating the effects of quality factors and organizational factors on university students' satisfaction of e-learning system quality ». Édité par Shuyan Wang. *Cogent Education* 7 (1): 1787004. https://doi.org/10.1080/2331186X.2020.1787004.

Alalwan, Nasser, Waleed Mugahed Al-Rahmi, Osama Alfarraj, Ahmed Alzahrani, Noraffandy Yahaya, et Ali Mugahed Al-Rahmi. 2019. « Integrated Three Theories to Develop a Model of Factors Affecting Students' Academic Performance in Higher Education ». *IEEE Access* 7: 98725-42. https://doi.org/10.1109/ACCESS.2019.2928142.

Alami, Nabil, Mohammed Meknassi, Noureddine En-nahnahi, Yassine El Adlouni, et Ouafae Ammor. 2021. « Unsupervised neural networks for automatic Arabic text summarization using document clustering and topic modeling ». *Expert Systems with Applications* 172: 114652.

Alaoui, Sara, Younès EL Bouzekri EL Idrissi, et Rachida Ajhoun. 2015. « Building Rich User Profile Based on Intentional Perspective ». *Procedia Computer Science*, International Conference on Advanced Wireless Information and Communication Technologies (AWICT 2015), 73 (janvier): 342-49. https://doi.org/10.1016/j.procs.2015.12.002.

Aldowah, Hanan, Hosam Al-Samarraie, Ahmed Ibrahim Alzahrani, et Nasser Alalwan. 2019. « Factors affecting student dropout in MOOCs: a cause and effect decision-making model ». *Journal of Computing in Higher Education*, 1-26.

Al-Fraihat, Dimah, Mike Joy, et Jane Sinclair. 2020. « Evaluating E-learning systems success: An empirical study ». *Computers in human behavior* 102: 67-86.

Alghizzawi, Mahmoud, Mohammed Habes, Said A Salloum, MA Ghani, C Mhamdi, et Khaled Shaalan. 2019. « The effect of social media usage on students'e-learning acceptance in higher education: A case study from the United Arab Emirates ». *Int. J. Inf. Technol. Lang. Stud* 3 (3): 13-26.

Alhijawi, Bushra, et Yousef Kilani. 2020. « A Collaborative Filtering Recommender System Using Genetic Algorithm ». *Information Processing & Management* 57 (6): 102310. https://doi.org/10.1016/j.ipm.2020.102310.

Al-Imarah, Ahmed A., et Robin Shields. 2019. « MOOCs, disruptive innovation and the future of higher education: A conceptual analysis ». *Innovations in Education and Teaching International* 56 (3): 258-69. https://doi.org/10.1080/14703297.2018.1443828.

Allen, I Elaine, et Jeff Seaman. 2013. *Changing course: Ten years of tracking online education in the United States*. ERIC.

Almaiah, Mohammed Amin, et Ahmed Al Mulhem. 2019. « Analysis of the Essential Factors Affecting of Intention to Use of Mobile Learning Applications: A Comparison between Universities Adopters and Non-Adopters ». *Education and Information Technologies* 24 (2): 1433-68. https://doi.org/10.1007/s10639-018-9840-1.

Alqurashi, Emtinan. 2016. « Self-Efficacy In Online Learning Environments: A Literature Review ». *Contemporary Issues in Education Research (CIER)* 9 (1): 45-52. https://doi.org/10.19030/cier.v9i1.9549.

Alsaeedi, Abdullah, et Mohammad Zubair Khan. 2019. « A study on sentiment analysis techniques of Twitter data ». *International Journal of Advanced Computer Science and Applications* 10 (2).

Alshammari, Sultan, Mohamad Bilal Ali, et Rosli M.S. 2018. « LMS, CMS AND LCMS: THE CONFUSION AMONG THEM » 30 (juin): 455-59.

Al-Shawabkeh, Abdallah, et Andriew Lim. 2014. « The Use of Social Media In Higher Education Learning: Swot Analysis of Using Social Media for Learning ». *ECSM 2014 University of Brighton Brighton, UK 10-11 July 2014*, 10.

Al-Yahya, Maha, Remya George, et Auhood Alfaries. 2015. « Ontologies in E-learning: review of the literature ». *International Journal of Software Engineering and Its Applications* 9 (2): 67-84.

Amane, Meryem, Karima Aissaoui, et Mohammed Berrada. 2022. « ERSDO: E-Learning Recommender System Based on Dynamic Ontology ». *Education and Information Technologies* 27 (6): 7549-61. https://doi.org/10.1007/s10639-022-10914-y.

Amarneh, Batool M., Muhammad Turki Alshurideh, Barween H. Al Kurdi, et Zaid Obeidat. 2021. « The Impact of COVID-19 on E-Learning: Advantages and Challenges ». In *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021)*, édité par Aboul Ella Hassanien, Abdelkrim Haqiq, Peter J. Tonellato, Ladjel Bellatreche, Sam Goundar, Ahmad Taher Azar, Essaid Sabir, et Driss Bouzidi, 75-89. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-76346-6_8.

Ameen, Ayesha, Khaleel Ur Rahman Khan, et B. Padmaja Rani. 2012. « Ontological student profile ». In *Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology*, 466-71. CCSEIT '12. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2393216.2393294.

Arun Gaikwad, VS. 2016. « E-Learning in India: wheel of change ». *International Journal of e-Education, e-Business, e-Management and e-Learning* 6 (0): 1.

Assami, Sara, Najima Daoudi, et Rachida Ajhoun. 2023. « Learner Profile Enrichment and Semantic Modeling of Learning Actors for MOOC Recommendation ». In *Proceedings of the 8th International Conference on Advanced Intelligent Systems and Informatics 2022*, édité par Aboul Ella Hassanien, Václav Snášel, Mincong Tang, Tien-Wen Sung, et Kuo-Chi Chang, 753-67. Lecture Notes on Data Engineering and Communications Technologies. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-20601-6_62.

Badali, Mehdi, Javad Hatami, Seyyed Kazem Banihashem, Ebrahim Rahimi, Omid Noroozi, et Zahra Eslami. 2022. « The role of motivation in MOOCs' retention rates: a systematic literature review ». *Research and Practice in Technology Enhanced Learning* 17 (1): 1-20.

Bai, Xiwen, Xiunian Zhang, Kevin X Li, Yaoming Zhou, et Kum Fai Yuen. 2021. « Research topics and trends in the maritime transport: A structural topic model ». *Transport Policy* 102: 11-24.

Bandura, Albert, et Richard H Walters. 1977. *Social learning theory*. Vol. 1. Prentice-hall Englewood Cliffs, NJ.

Barker, Philip. 2002. « On being an online tutor ». *Innovations in Education and Teaching International* 39 (1): 3-13.

Baturay, Meltem Huri. 2015. « An Overview of the World of MOOCs ». *Procedia - Social and Behavioral Sciences*, International Conference on New Horizons in Education, INTE 2014, 25-27 June 2014, Paris, France, 174 (février): 427-33. https://doi.org/10.1016/j.sbspro.2015.01.685.

Bedi, Pradeep, S. B. Goyal, Anand Singh Rajawat, Rabindra Nath Shaw, et Ankush Ghosh. 2022. « A Framework for Personalizing Atypical Web Search Sessions with Concept-Based User Profiles Using Selective Machine Learning Techniques ». In *Advanced Computing and Intelligent Technologies*, édité par Monica Bianchini, Vincenzo Piuri, Sanjoy Das, et Rabindra Nath Shaw, 279-91. Lecture Notes in Networks and Systems. Singapore: Springer. https://doi.org/10.1007/978-981-16-2164-2_23.

Beel, Joeran, Bela Gipp, Stefan Langer, et Corinna Breitinger. 2016. « Research-Paper Recommender Systems: A Literature Survey ». *International Journal on Digital Libraries* 17 (4): 305-38. https://doi.org/10.1007/s00799-015-0156-0.

Behera, Santosh Kumar. 2013. « E-and M-Learning: A comparative study ». *International journal on new trends in education and their implications* 4 (3): 65-78.

Benhamdi, Soulef, Abdesselam Babouri, et Raja Chiky. 2017. « Personalized Recommender System for E-Learning Environment ». *Education and Information Technologies* 22 (4): 1455-77. https://doi.org/10.1007/s10639-016-9504-y.

Bennani, Souha, Ahmed Maalel, et Henda Ben Ghezala. 2020. « AGE-Learn: Ontology-based representation of personalized gamification in E-learning ». *Procedia Computer Science* 176: 1005-14.

Besel, Christoph, Jörg Schlötterer, et Michael Granitzer. 2016. « Inferring semantic interest profiles from Twitter followees: does Twitter know better than your friends? » In , 1152-57.

Bezerra, Luis N., Márcia T. Silva, AUTOR 3, AUTOR 4, AUTOR 5, et AUTOR 6. 2017. « A review of literature on the reasons that cause the high dropout rates in the MOOCS ». *Revista ESPACIOS* 38 (05). https://www.revistaespacios.com/a17v38n05/17380511.html#thed.

Bhattacharya, Parantapa, Muhammad Bilal Zafar, Niloy Ganguly, Saptarshi Ghosh, et Krishna P Gummadi. 2014. « Inferring user interests in the twitter social network ». In , 357-60.

Bianchi, Federico, Silvia Terragni, et Dirk Hovy. 2020. « Pre-training is a hot topic: Contextualized document embeddings improve topic coherence ». *arXiv preprint arXiv:2004.03974*.

Bilal, Muhammad, Abdullah Gani, Muhammad Ikram Ullah Lali, Mohsen Marjani, et Nadia Malik. 2019. « Social profiling: A review, taxonomy, and challenges ». *Cyberpsychology, Behavior, and Social Networking* 22 (7): 433-50.

Blei, David M, Andrew Y Ng, et Michael I Jordan. 2003. « Latent dirichlet allocation ». *Journal of machine Learning research* 3 (Jan): 993-1022.

Blevins, Terra, et Luke Zettlemoyer. 2020. « Moving Down the Long Tail of Word Sense Disambiguation with Gloss-Informed Biencoders ». arXiv. https://doi.org/10.48550/arXiv.2005.02590.

Bobillo, Fernando, et Umberto Straccia. 2016. « The fuzzy ontology reasoner fuzzyDL ». *Knowledge-Based Systems* 95: 12-34.

Bondarouk, Tanya, et Huub Ruël. 2010. « Dynamics of e-learning: theoretical and practical perspectives: Introduction to special issue ». *International journal of training and development* 14 (3): 149-54.

Boudchiche, Mohamed, et Azzeddine Mazroui. 2019. « A Hybrid Approach for Arabic Lemmatization ». *International Journal of Speech Technology* 22 (3): 563-73. https://doi.org/10.1007/s10772-018-9528-3.

Bouneffouf, Djallel. 2013. « Towards user profile modelling in recommender system ». *arXiv preprint arXiv:1305.1114*.

Bourekkache, Samir, Okba Kazar, Mounia Abik, Samir Tigane, et Laid Kahloul. 2019. « Ontology based approach for representing the learner profile and learning styles ». In

2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), 1-6. https://doi.org/10.1109/ICDS47004.2019.8942276.

Bourkoukou, Outmane, et Essaid El Bachari. 2018. « Toward a Hybrid Recommender System for E-Learning Personnalization Based on Data Mining Techniques ». *JOIV : International Journal on Informatics Visualization* 2 (4): 271-78. https://doi.org/10.30630/joiv.2.4.158.

Brady, Kevin P, Lori B Holcomb, et Bethany V Smith. 2010. « The use of alternative social networking sites in higher educational settings: A case study of the e-learning benefits of Ning in education. » *Journal of interactive online learning* 9 (2).

Briggs Myers, Isabel, et Peter B Myers. 1995. « Gifts differing: Understanding personality type ». *Davis-Black Publishing*.

BURKE, ROBIN. 2002. « Hybrid Recommender Systems: Survey and Experiments », 40.

Cabot, Jordi. 2018. « WordPress: A Content Management System to Democratize Publishing ». *IEEE Software* 35 (3): 89-92. https://doi.org/10.1109/MS.2018.2141016.

Calegari, Silvia, et Gabriella Pasi. 2013. « Personal Ontologies: Generation of User Profiles Based on the YAGO Ontology ». *Information Processing & Management* 49 (3): 640-58. https://doi.org/10.1016/j.ipm.2012.07.010.

Calero, Coral, Francisco Ruiz, et Mario Piattini. 2006. *Ontologies for software engineering and software technology*. Springer Science & Business Media.

Castellanos-Reyes, Daniela, Yukiko Maeda, et Jennifer C Richardson. 2021. « THE RELATIONSHIP BETWEEN SOCIAL NETWORK SITES AND PERCEIVED LEARNING AND SATISFACTION FOR EDUCATIONAL PURPOSES ». *Social Media: Influences on Education*, 231.

Chatterjee, Niladri, Neha Kaushik, Deepali Gupta, et Ramneek Bhatia. 2017. « Ontology merging: A practical perspective ». In , 136-45. Springer.

Chen, Chaotao, et Jiangtao Ren. 2017. « Forum latent Dirichlet allocation for user interest discovery ». *Knowledge-Based Systems* 126: 1-7.

Chen, Jing, Jun Feng, Xia Sun, Nannan Wu, Zhengzheng Yang, et Sushing Chen. 2019. « MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine ». *Mathematical Problems in Engineering* 2019.

Chicaiza, Janneth, Nelson Piedra, Jorge Lopez-Vargas, et Edmundo Tovar-Caro. 2015. « A user profile definition in context of recommendation of open educational resources. An approach based on linked open vocabularies ». In *Frontiers in Education Conference (FIE), 2015 IEEE*, 1-7. IEEE.

Choi, Namyoun, Il-Yeol Song, et Hyoil Han. 2006. « A survey on ontology mapping ». *ACM Sigmod Record* 35 (3): 34-41.

Choudhury, Snigdha, et Snigdha Pattnaik. 2020. « Emerging Themes in E-Learning: A Review from the Stakeholders' Perspective ». *Computers & Education* 144 (janvier): 103657. https://doi.org/10.1016/j.compedu.2019.103657.

Chrysafiadi, Konstantina, et Maria Virvou. 2013. « Student Modeling Approaches: A Literature Review for the Last Decade ». *Expert Systems with Applications* 40 (11): 4715-29. https://doi.org/10.1016/j.eswa.2013.02.007.

Chrysafiadi, K., & Virvou, M. 2015. « Fuzzy Logic for Adaptive Instruction in an E-learning Environment for Computer Programming ». *IEEE Transactions on Fuzzy Systems* 23 (1): 164-77. https://doi.org/10.1109/TFUZZ.2014.2310242.

Chukwuere, Joshua Ebere, et Precious Chibuike Chukwuere. 2017. « The impact of social media on social lifestyle: A case study of university female students ». *Gender and Behaviour* 15 (4): 9966-81.

Cidral, Wilmar Audye, Tiago Oliveira, Massimo Di Felice, et Manuela Aparicio. 2018a. « E-learning success determinants: Brazilian empirical study ». *Computers & Education* 122: 273-90.

Cidral, W. A., Oliveira, T., Di Felice, M., & Aparicio, M. 2018b. « E-Learning Success Determinants: Brazilian Empirical Study ». *Computers & Education* 122 (juillet): 273-90. https://doi.org/10.1016/j.compedu.2017.12.001.

Coleman, Barbara Carrick, Stacie K. Pettit, et Megan M. Buning. 2018. « Social Media Use in Higher Education: Do Members of the Academy Recognize Any Advantages? » *The Journal of Social Media in Society* 7 (1): 420-42.

Corcho, Oscar, Mariano Fernández-López, et Asunción Gómez-Pérez. 2003. « Methodologies, Tools and Languages for Building Ontologies. Where Is Their Meeting Point? » *Data & Knowledge Engineering* 46 (1): 41-64. https://doi.org/10.1016/S0169-023X(02)00195-7.

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. 2006. « Ontological engineering: principles, methods, tools and languages ». In *Ontologies for software engineering and software technology*, 1-48. Springer.

Corcho, Oscar, Mariano Fernández-López, Asunción Gómez-Pérez, et Angel López-Cima. 2005. « Building legal ontologies with METHONTOLOGY and WebODE ». In *Law and the semantic web*, 142-57. Springer.

Corcho, Oscar, María Poveda-Villalón, et Asunción Gómez-Pérez. 2015. « Ontology engineering in the era of linked data ». *Bulletin of the Association for Information Science and Technology* 41 (4): 13-17.

Crane, Rich A, et Steph Comley. 2021. « Influence of social learning on the completion rate of massive online open courses ». *Education and Information Technologies* 26 (2): 2285-93.

Cufoglu, Ayse. 2014. « User profiling-a short review ». *International Journal of Computer Applications* 108 (3).

D. Nawara et R. Kashef. 2021. « Context-Aware Recommendation Systems in the IoT Environment (IoT-CARS)–A Comprehensive Overview ». *IEEE Access* 9: 144270-84. https://doi.org/10.1109/ACCESS.2021.3122098.

Dahdouh, Karim, Ahmed Dakkak, Lahcen Oughdir, et Abdelali Ibriz. 2019. « Large-Scale e-Learning Recommender System Based on Spark and Hadoop ». *Journal of Big Data* 6 (1): 2. https://doi.org/10.1186/s40537-019-0169-4.

Dan, W. U., et WANG Hui-lin. 2006. « Role of ontology in information retrieval ». *Journal of Electronic Science and Technology* 4 (2): 148-54.

Dass, Sheran, Kevin Gary, et James Cunningham. 2021. « Predicting Student Dropout in Self-Paced MOOC Course Using Random Forest Model ». *Information* 12 (11): 476. https://doi.org/10.3390/info12110476.

De Cubber, G. 2001. « The killer application: E-learning terminology, components, adverse factors,(r) e-volution ». *Business Quest*.

Deaton, Shannon. 2015. « Social Learning Theory in the Age of Social Media: Implications for Educational Practitioners ». *Journal of Educational Technology* 12 (1): 1-6.

Deebak, B. D., et Fadi Al-Turjman. 2020. « A novel community-based trust aware recommender systems for big data cloud service networks ». *Sustainable Cities and Society* 61 (octobre): 102274. https://doi.org/10.1016/j.scs.2020.102274.

Deng, Ruiqi, Pierre Benckendorff, et Deanne Gannaway. 2019. « Progress and New Directions for Teaching and Learning in MOOCs ». *Computers & Education* 129 (février): 48-60. https://doi.org/10.1016/j.compedu.2018.10.019.

Dichy, Joseph, et Ali Farghaly. 2003. « Roots & patterns vs. stems plus grammar-lexis specifications: On what basis should a multilingual database centred on Arabic be built? » In .

Diera, Andor, Bao Xin Lin, Bhakti Khera, Tim Meuser, Tushar Singhal, Lukas Galke, et Ansgar Scherp. 2022. « Bag-of-Words vs. Sequence vs. Graph vs. Hierarchy for Single-and Multi-Label Text Classification ». *arXiv preprint arXiv:2204.03954*.

« Digital 2021 April Statshot Report ». s. d. DataReportal – Global Digital Insights. Consulté le 26 novembre 2022. https://datareportal.com/reports/digital-2021-april-global-statshot.

Do, Hong-Hai, et Erhard Rahm. 2002. « COMA—a system for flexible combination of schema matching approaches ». In *VLDB'02: Proceedings of the 28th International Conference on Very Large Databases*, 610-21. Elsevier.

Domingue, John. 1998. « Tadzebao and WebOnto: Discussing, browsing, and editing ontologies on the web ».

Dominici, Gandolfo, et Federica Palumbo. 2013. « How to Build an E-Learning Product: Factors for Student/Customer Satisfaction ». *Business Horizons* 56 (1): 87-96. https://doi.org/10.1016/j.bushor.2012.09.011.

Downes, Stephen. 2005. « E-learning 2.0. » *Elearn magazine* 2005 (10): 1.

Dron, Jon, et Terry Anderson. 2016. « The Future of E-learning ». In .

Duong, Trong Hai, Ngoc Thanh Nguyen, et Geun Sik Jo. 2010. « Constructing and mining a semantic-based academic social network ». *Journal of Intelligent & Fuzzy Systems* 21 (3): 197-207. https://doi.org/10.3233/IFS-2010-0451.

Duong, Trong Hai, Mohammed Nazim Uddin, Delong Li, et Geun Sik Jo. 2009. « A Collaborative Ontology-Based User Profiles System ». In *Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems*, 540-52. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04441-0_47.

Dwivedi, P., et K.K. Bharadwaj. 2015. « E-Learning recommender system for a group of learners based on the unified learner profile approach ». *Expert Systems* 32 (2): 264-76. https://doi.org/10.1111/exsy.12061.

Dzisevič, Robert, et Dmitrij Šešok. 2019. « Text Classification using Different Feature Extraction Approaches ». In *2019 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, 1-4. https://doi.org/10.1109/eStream.2019.8732167.

Echeng, Razep, Abel Usoro, et Isaiah Ewuzie. 2016. « Factors to Consider when Enhancing the Use of Web 2.0 Technologies in Higher Education: Students' and Lectures' Views for Quality Use ». *International Journal for Digital Society* 7 (mars). https://doi.org/10.20533/ijds.2040.2570.2016.0138.

Eke, Christopher Ifeanyi, Azah Anir Norman, Liyana Shuib, et Henry Friday Nweke. 2019. « A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions ». *IEEE Access* 7: 144907-24. https://doi.org/10.1109/ACCESS.2019.2944243.

Ekwonwune, Emmanuel N., et Dominic C. Edebatu. 2019. « Design and Implementation of an Online Course Management System ». *Journal of Software Engineering and Applications* 12 (2): 21-33. https://doi.org/10.4236/jsea.2019.122002.

EL AISSAOUI, Ouafae, et Lahcen OUGHDIR. 2020. « A learning style-based Ontology Matching to enhance learning resources recommendation ». In , 1-7. IEEE.

El Said, Ghada Refaat. 2017. « Understanding how learners use massive open online courses and why they drop out: Thematic analysis of an interview study in a developing country ». *Journal of Educational Computing Research* 55 (5): 724-52.

Elallioui, Youssouf, et Omar El Beqqali. 2012. « User profile Ontology for the Personalization approach ». *International Journal of Computer Applications* 41 (mars): 31-40. https://doi.org/10.5120/5531-7577.

El-Bishouty, Moushir M., Ahmed Aldraiweesh, Uthman Alturki, Richard Tortorella, Junfeng Yang, Ting-Wen Chang, Sabine Graf, et Kinshuk. 2019. « Use of Felder and Silverman Learning Style Model for Online Course Design ». *Educational Technology Research and Development* 67 (1): 161-77. https://doi.org/10.1007/s11423-018-9634-6.

Elias, Mirette, Steffen Lohmann, et Sören Auer. 2017. « Ontology-Based Representation of Learner Profiles for Accessible OpenCourseWare Systems ». In *Knowledge Engineering and Semantic Web*, édité par Przemysław Różewski et Christoph Lange, 279-94. Communications in Computer and Information Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-69548-8_19.

Eriksson, Thommy, Tom Adawi, et Christian Stöhr. 2017. « "Time Is the Bottleneck": A Qualitative Study Exploring Why Learners Drop out of MOOCs ». *Journal of Computing in Higher Education* 29 (1): 133-46. https://doi.org/10.1007/s12528-016-9127-8.

Er-Rafyg, Aicha, Manar Abourezq, Abdellah Idrissi, et Adil Bouhouch. s. d. « Courses Recommendations Using Skyline BNL Algorithm », 19.

Eyharabide, Victoria, et Analía Amandi. 2012. « Ontology-Based User Profile Learning ». *Applied Intelligence* 36 (4): 857-69. https://doi.org/10.1007/s10489-011-0301-4.

Farquhar, Adam, Richard Fikes, et James Rice. 1997. « The ontolingua server: A tool for collaborative ontology construction ». *International journal of human-computer studies* 46 (6): 707-27.

Farseev, Aleksandr, Liqiang Nie, Mohammad Akbari, et Tat-Seng Chua. 2015. « Harvesting multiple sources for user profile learning: a big data study ». In , 235-42.

Fayyaz, Zeshan, Mahsa Ebrahimian, Dina Nawara, Ahmed Ibrahim, et Rasha Kashef. 2020. « Recommendation systems: Algorithms, challenges, metrics, and business opportunities ». *applied sciences* 10 (21): 7748.

Felder, Richard M. 2002. « Learning and teaching styles in engineering education ».

Feng, Wenzheng, Jie Tang, et Tracy Xiao Liu. 2019. « Understanding dropouts in MOOCs ». In , 33:517-24.

FernáNdez-LóPez, Mariano, et AsuncióN GóMez-PéRez. 2002. « Overview and Analysis of Methodologies for Building Ontologies ». *The Knowledge Engineering Review* 17 (02). https://doi.org/10.1017/S0269888902000462.

Fernández-López, Mariano, Asunción Gómez-Pérez, et Natalia Juristo. 1997a. « Methontology: from ontological art towards ontological engineering ».

———. 1997b. « Methontology: from ontological art towards ontological engineering ».

« FOAF Vocabulary Specification ». s. d. Consulté le 11 février 2020. http://xmlns.com/foaf/spec/.

Frias-Martinez, E., G. Magoulas, S. Chen, et R. Macredie. 2005. « Modeling Human Behavior in User-Adaptive Systems: Recent Advances Using Soft Computing Techniques ». *Expert Systems with Applications* 29 (2): 320-29. https://doi.org/10.1016/j.eswa.2005.04.005.

Gao, Wei, Mohammad Reza Farahani, Adnan Aslam, et Sunilkumar Hosamani. 2017. « Distance Learning Techniques for Ontology Similarity Measuring and Ontology Mapping ». *Cluster Computing* 20 (2): 959-68. https://doi.org/10.1007/s10586-017-0887-3.

García, Salvador, Julián Luengo, et Francisco Herrera. 2015. *Data Preprocessing in Data Mining*. Vol. 72. Intelligent Systems Reference Library. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10247-4.

García-Peñalvo, Francisco J., Ángel Fidalgo-Blanco, et María Luisa Sein-Echaluce. 2018. « An Adaptive Hybrid MOOC Model: Disrupting the MOOC Concept in Higher Education ». *Telematics and Informatics* 35 (4): 1018-30. https://doi.org/10.1016/j.tele.2017.09.012.

Garrison, D Randy. 2016. *E-learning in the 21st century: A community of inquiry framework for research and practice*. Routledge.

Ghenname, Mérième. 2015. « Le web social et le web sémantique pour la recommandation de ressources pédagogiques ».

Gherheş, Vasile, Claudia E. Stoian, Marcela Alina Fărcașiu, et Miroslav Stanici. 2021. « E-Learning vs. Face-To-Face Learning: Analyzing Students' Preferences and Behaviors ». *Sustainability* 13 (8): 4381. https://doi.org/10.3390/su13084381.

Ghosh, Riddhiman. 2016. System and method for generating a user profile, issued mars 2016.

Giannakos, Michail N., Patrick Mikalef, et Ilias O. Pappas. 2022. « Systematic Literature Review of E-Learning Capabilities to Enhance Organizational Learning ». *Information Systems Frontiers* 24 (2): 619-35. https://doi.org/10.1007/s10796-020-10097-2.

Gilbert, Patricia K., et Nada Dabbagh. 2005. « How to Structure Online Discussions for Meaningful Discourse: A Case Study ». *British Journal of Educational Technology* 36 (1): 5-18. https://doi.org/10.1111/j.1467-8535.2005.00434.x.

Giri, Ritwik, Heesook Choi, Kevin Soo Hoo, et Bhaskar D Rao. 2014. « User behavior modeling in a cellular network using latent dirichlet allocation ». In , 36-44. Springer.

Godoy, Daniela, et Analia Amandi. 2005. « User profiling in personal information agents: a survey ». *The Knowledge Engineering Review* 20 (4): 329-61.

Goldberg, David, David Nichols, Brian M. Oki, et Douglas Terry. 1992. « Using Collaborative Filtering to Weave an Information Tapestry ». *Communications of the ACM* 35 (12): 61-70. https://doi.org/10.1145/138859.138867.

Golemati, Maria, Akrivi Katifori, Costas Vassilakis, George Lepouras, et Constantin Halatsis. 2007. « Creating an ontology for the user profile: Method and applications ». In *Proceedings of the first RCIS conference*, 407-12.

Gómez-Pérez, Asunción, Mariano Fernández-López, et Oscar Corcho. 2004. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web / Asunción Gómez-Pérez, Mariano Fernández-López, and Oscar Corcho*. Advanced information and knowledge processing. London ; New York: Springer.

Gómez-Pérez, A., Fernández-López, M., & Corcho, O.2006. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media.

Gomez-Zermeno, Marcela Gerogina, et Lorena Aleman de La Garza. 2016. « Research analysis on MOOC course dropout and retention rates ». *Turkish online journal of distance education* 17 (2).

Goopio, Joselyn, et Catherine Cheung. 2021. « The MOOC dropout phenomenon and retention strategies ». *Journal of Teaching in Travel & Tourism* 21 (2): 177-97. https://doi.org/10.1080/15313220.2020.1809050.

Govindasamy, Thavamalar. 2001. « Successful Implementation of E-Learning: Pedagogical Considerations ». *The Internet and Higher Education*, SOLE 2001: Exploring the Evolution of E-Learning, 4 (3): 287-99. https://doi.org/10.1016/S1096-7516(01)00071-9.

Goyal, Archana, Vishal Gupta, et Manish Kumar. 2018. « Recent Named Entity Recognition and Classification Techniques: A Systematic Review ». *Computer Science Review* 29 (août): 21-43. https://doi.org/10.1016/j.cosrev.2018.06.001.

Graves, Mike, Adam Constabaris, et Dan Brickley. 2007. « Foaf: Connecting people on the semantic web ». *Cataloging & classification quarterly* 43 (3-4): 191-202.

Gray, Audrey Joyce. 1997. *Constructivist teaching and learning*. Saskatchewan School Trustees Association.

Greenhow, Christine, et Emilia Askari. 2017. « Learning and teaching with social network sites: A decade of research in K-12 related education ». *Education and information technologies* 22 (2): 623-45.

Grimm, Stephan, Andreas Abecker, Johanna Völker, et Rudi Studer. 2011. « Ontologies and the Semantic Web ». In *Handbook of Semantic Web Technologies*, édité par John Domingue, Dieter Fensel, et James A. Hendler, 507-79. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-92913-0_13.

Grootendorst, Maarten. 2020. « Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics ». *URL https://doi. org/10.5281/zenodo* 4381785.

Grootendorst, M. 2022. « BERTopic: Neural topic modeling with a class-based TF-IDF procedure ». *arXiv preprint arXiv:2203.05794*.

Gruber, Thomas R. 1995. « Toward principles for the design of ontologies used for knowledge sharing? » *International journal of human-computer studies* 43 (5-6): 907-28.

References

Guarino, Nicola. 1995. « Formal ontology, conceptual analysis and knowledge representation ». *International journal of human-computer studies* 43 (5-6): 625-40.

Guo, Chonghui, Menglin Lu, et Wei Wei. 2021. « An improved LDA topic modeling method based on partition for medium and long texts ». *Annals of Data Science* 8 (2): 331-44.

Gupta, Aakansha, et Rahul Katarya. 2021. « PAN-LDA: A latent Dirichlet allocation based novel feature extraction model for COVID-19 data using machine learning ». *Computers in biology and medicine* 138: 104920.

Gupta, Kriti Priya, et Harshit Maurya. 2022. « Adoption, completion and continuance of MOOCs: a longitudinal study of students' behavioural intentions ». *Behaviour & Information Technology* 41 (3): 611-28.

Gupta, Rahul Kumar, Ritu Agarwalla, Bukya Hemanth Naik, Joythish Reddy Evuri, Apil Thapa, et Thoudam Doren Singh. 2022. « Prediction of Research Trends Using LDA Based Topic Modeling ». *Global Transitions Proceedings*, International Conference on Intelligent Engineering Approach(ICIEA-2022), 3 (1): 298-304. https://doi.org/10.1016/j.gltp.2022.03.015.

Gurl, Emet. 2017. « SWOT ANALYSIS: A THEORETICAL REVIEW », août. http://hdl.handle.net/10673/792.

Guruge, Deepani B., Rajan Kadel, et Sharly J. Halder. 2021. « The State of the Art in Methodologies of Course Recommender Systems—A Review of Recent Research ». *Data* 6 (2): 18. https://doi.org/10.3390/data6020018.

Gütl, Christian, Rocael Hernández Rizzardini, Vanessa Chang, et Miguel Morales. 2014. « Attrition in MOOC: Lessons learned from drop-out students ». In , 37-48. Springer.

Haider, Mofiz Mojib, Md Arman Hossin, Hasibur Rashid Mahi, et Hossain Arif. 2020. « Automatic text summarization using gensim word2vec and k-means clustering algorithm ». In , 283-86. IEEE.

Hajli, Mahmood, Hatem Bugshan, Xiaolin Lin, et Mauricio Featherman. 2013. « From e-learning to social learning – a health care study ». *European Journal of Training and Development* 37 (9): 851-63. https://doi.org/10.1108/EJTD-10-2012-0062.

Hammouda, Faten Khalfallah, et Abdelsalam Abdelhamid Almarimi. 2010. « Heuristic lemmatization for Arabic texts indexation and classification 1 ».

Han, Jiying, Xiaohui Geng, et Qinxiang Wang. 2021. « Sustainable Development of University EFL Learners' Engagement, Satisfaction, and Self-Efficacy in Online Learning Environments: Chinese Experiences ». *Sustainability* 13 (21): 11655. https://doi.org/10.3390/su132111655.

Han, Lixin, et Guihai Chen. 2009. « A Fuzzy Clustering Method of Construction of Ontology-Based User Profiles ». *Advances in Engineering Software* 40 (7): 535-40. https://doi.org/10.1016/j.advengsoft.2008.10.006.

Harackiewicz, Judith M, Jessi L Smith, et Stacy J Priniski. 2016. « Interest matters: The importance of promoting interest in education ». *Policy insights from the behavioral and brain sciences* 3 (2): 220-27.

Hardy, Bruce W, et Jessica Castonguay. 2018. « The moderating role of age in the relationship between social media use and mental well-being: An analysis of the 2016 General Social Survey ». *Computers in Human Behavior* 85: 282-90.

Harrow, Ian, Rama Balakrishnan, Ernesto Jimenez-Ruiz, Simon Jupp, Jane Lomax, Jane Reed, Martin Romacker, et al. 2019. « Ontology Mapping for Semantically Enabled Applications ». *Drug Discovery Today* 24 (10): 2068-75. https://doi.org/10.1016/j.drudis.2019.05.020.

Hasan, Md. Rakibul, Maisha Maliha, et M. Arifuzzaman. 2019. « Sentiment Analysis with NLP on Twitter Data ». In *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, 1-4. https://doi.org/10.1109/IC4ME247184.2019.9036670.

He, Jun, Hongyan Liu, Yiqing Zheng, Shu Tang, Wei He, et Xiaoyong Du. 2020. « Bi-labeled LDA: Inferring interest tags for non-famous users in social network ». *Data Science and Engineering* 5 (1): 27-47.

Heidari, Maryam, et James H Jones. 2020. « Using bert to extract topic-independent sentiment features for social media bot detection ». In , 0542-47. IEEE.

Hoblitzell, Andrew, Meghna Babbar-Sebens, et Snehasis Mukhopadhyay. 2016. « Fuzzy and deep learning approaches for user modeling in wetland design ». In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 002133-38. https://doi.org/10.1109/SMC.2016.7844555.

Horsburgh, Jo, et Kate Ippolito. 2018. « A skill to be worked at: using social learning theory to explore the process of learning from role models in clinical settings ». *BMC Medical Education* 18 (1): 156. https://doi.org/10.1186/s12909-018-1251-x.

Horvitz, Eric J., John S. Breese, David Heckerman, David Hovel, et Koos Rommelse. 2013. « The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users ». arXiv. https://doi.org/10.48550/arXiv.1301.7385.

Hosen, Mosharrof, Samuel Ogbeibu, Beena Giridharan, Tat-Huei Cham, Weng Marc Lim, et Justin Paul. 2021. « Individual Motivation and Social Media Influence on Student Knowledge Sharing and Learning Performance: Evidence from an Emerging Economy ». *Computers & Education* 172 (octobre): 104262. https://doi.org/10.1016/j.compedu.2021.104262.

Howard, Richard A, Curtis A Carver, et William D Lane. 1996. « Felder's learning styles, Bloom's taxonomy, and the Kolb learning cycle: tying it all together in the CS2 course ». In , 227-31.

Iliadis, Andrew, Amelia Acker, Wesley Stevens, et Sezgi Başak Kavakli. 2022. « One schema to rule them all: How Schema. org models the world of search ». *Journal of the Association for Information Science and Technology*.

Imran, Hazra, Mohammad Belghis-Zadeh, Ting-Wen Chang, Kinshuk, et Sabine Graf. 2016. « PLORS: A Personalized Learning Object Recommender System ». *Vietnam Journal of Computer Science* 3 (1): 3-13. https://doi.org/10.1007/s40595-015-0049-6.

« IMS Learner Information Packaging Information Model Specification | IMS Global Learning Consortium ». s. d. Consulté le 11 février 2020. https://www.imsglobal.org/profiles/lipinfo01.html.

Itani, Alya, Laurent Brisson, et Serge Garlatti. 2018. « Understanding learner's drop-out in MOOCs ». In , 233-44. Springer.

Jannach, Dietmar, et Michael Jugovac. 2019. « Measuring the business value of recommender systems ». *ACM Transactions on Management Information Systems (TMIS)* 10 (4): 1-23.

Jeevamol, Joy, et V. G. Renumol. 2021. « An Ontology-Based Hybrid e-Learning Content Recommender System for Alleviating the Cold-Start Problem ». *Education and Information Technologies* 26 (4): 4993-5022. https://doi.org/10.1007/s10639-021-10508-0.

Jena, Kalyan Kumar, Sourav Kumar Bhoi, Tushar Kanta Malik, Kshira Sagar Sahoo, N. Z. Jhanjhi, Sajal Bhatia, et Fathi Amsaad. 2023. « E-Learning Course Recommender System Using Collaborative Filtering Models ». *Electronics* 12 (1): 157. https://doi.org/10.3390/electronics12010157.

Ji, Hyangeun, Soyeon Park, et Hye Won Shin. 2022. « Investigating the Link between Engagement, Readiness, and Satisfaction in a Synchronous Online Second Language Learning Environment ». *System* 105 (avril): 102720. https://doi.org/10.1016/j.system.2022.102720.

Jin, Cong. 2021. « Dropout Prediction Model in MOOC Based on Clickstream Data and Student Sample Weight ». *Soft Computing* 25 (14): 8971-88. https://doi.org/10.1007/s00500-021-05795-1.

Joy, Jeevamol, et Nisha S Raj. 2019. « An ontology model for content recommendation in personalized learning environment ». In , 1-6.

Joy, Jeevamol, Nisha S. Raj, et Renumol V. G. 2021. « Ontology-based E-learning Content Recommender System for Addressing the Pure Cold-start Problem ». *Journal of Data and Information Quality* 13 (3): 16:1-16:27. https://doi.org/10.1145/3429251.

Jung, Sooyoung, et Jun-Ho Huh. 2019. « An Efficient LMS Platform and Its Test Bed ». *Electronics* 8 (2): 154. https://doi.org/10.3390/electronics8020154.

Kaladevi Ramar et Geetha Gurunathan. s. d. « Technical review on ontology mapping techniques ».

Kang, Yue, Zhao Cai, Chee-Wee Tan, Qian Huang, et Hefu Liu. 2020. « Natural language processing (NLP) in management research: A literature review ». *Journal of Management Analytics* 7 (2): 139-72. https://doi.org/10.1080/23270012.2020.1756939.

Kanoje, Sumitkumar, Sheetal Girase, et Debajyoti Mukhopadhyay. 2015. « User Profiling Trends, Techniques and Applications ». arXiv. https://doi.org/10.48550/arXiv.1503.07474.

Kanoje, Sumitkumar, Debajyoti Mukhopadhyay, et Sheetal Girase. 2016. « User Profiling for University Recommender System Using Automatic Information Retrieval ». *Procedia Computer Science*, 1st International Conference on Information Security & Privacy 2015, 78 (janvier): 5-12. https://doi.org/10.1016/j.procs.2016.02.002.

Kapoor, Bhaskar, et Savita Sharma. 2010. « A comparative study ontology building tools for semantic web applications ». *International Journal of Web & Semantic Technology (IJWesT)* 1 (3): 1-13.

Kattoua, Tagreed, et Musa Al-Lozi. 2016. « A Review of Literature on E-Learning Systems in Higher Education » 7: 9.

Kavcic, A. 2004. « Fuzzy user modeling for adaptation in educational hypermedia ». *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34 (4): 439-49. https://doi.org/10.1109/TSMCC.2004.833294.

Keefe, James W, et Owen B Kiernan. 1979. *Student learning styles: Diagnosing and prescribing programs*. National Assn of Secondary School.

Khadir, Ahlem Chérifa, Hassina Aliane, et Ahmed Guessoum. 2021. « Ontology Learning: Grand Tour and Challenges ». *Computer Science Review* 39 (février): 100339. https://doi.org/10.1016/j.cosrev.2020.100339.

Khalil, Hanan, et Martin Ebner. 2014. « MOOCs completion rates and possible methods to improve retention-A literature review ». *EdMedia+ innovate learning*, 1305-13.

Khanal, Shristi Shakya, PWC Prasad, Abeer Alsadoon, et Angelika Maag. 2020. « A systematic review: machine learning based recommendation systems for e-learning ». *Education and Information Technologies* 25: 2635-64.

Kim, Jonghun, Daesung Lee, et Kyung-Yong Chung. 2014. « Item recommendation based on context-aware model for personalized u-healthcare service ». *Multimedia Tools and Applications* 71 (2): 855-72. https://doi.org/10.1007/s11042-011-0920-0.

Kim, Kunyoung, Jongmo Kim, Minhwan Kim, et Mye Sohn. 2021. « User interest-based recommender system for image-sharing social media ». *World Wide Web* 24 (3): 1003-25.

Kim, Seong-kyu, et Jun-Ho Huh. 2019. « A Study on the LMS Platform Performance and Performance Improvement of K-MOOCs Platform from Learner's Perspective ». In *Advanced Multimedia and Ubiquitous Engineering*, édité par James J. Park, Vincenzo Loia, Kim-Kwang Raymond Choo, et Gangman Yi, 781-86. Lecture Notes in Electrical Engineering. Singapore: Springer. https://doi.org/10.1007/978-981-13-1328-8_101.

Kim, Sin-Hyang, et Sihyun Park. 2021. « Influence of Learning Flow and Distance E-Learning Satisfaction on Learning Outcomes and the Moderated Mediation Effect of Social-Evaluative Anxiety in Nursing College Students during the COVID-19 Pandemic: A Cross-Sectional Study ». *Nurse Education in Practice* 56 (octobre): 103197. https://doi.org/10.1016/j.nepr.2021.103197.

Kim, Younghoon, et Kyuseok Shim. 2014. « TWILITE: A Recommendation System for Twitter Using a Probabilistic Model Based on Latent Dirichlet Allocation ». *Information Systems* 42 (juin): 59-77. https://doi.org/10.1016/j.is.2013.11.003.

Klašnja-Milićević, Aleksandra, Boban Vesin, Mirjana Ivanović, et Zoran Budimac. 2011. « E-Learning personalization based on hybrid recommendation strategy and learning style identification ». *Computers & Education* 56 (3): 885-99.

Klobas, Jane E., et Tanya J. McGill. 2010. « The Role of Involvement in Learning Management System Success ». *Journal of Computing in Higher Education* 22 (2): 114-34. https://doi.org/10.1007/s12528-010-9032-5.

Kolokytha, Eleftheria, Sofia Loutrouki, Stavros Valsamidis, et Giannoula Florou. 2015. « Social media networks as a learning tool ». *Procedia Economics and Finance* 19: 287-95.

Kordahi, Marilou. 2022. « Ontology for the User-Learner Profile Personalizes the Search Analysis of Online Learning Resources ». *Information Technology and Libraries* 41 (2).

Koren, Yehuda, Steffen Rendle, et Robert Bell. 2021. « Advances in collaborative filtering ». *Recommender systems handbook*, 91-142.

Kourtiche, Ali, Sidi mohamed Benslimane, et Sofiane Boukli Hacene. 2020. « OUPIP: Ontology Based User Profile for Impairment Person in Dynamic Situation Aware Social Networks ». *International Journal of Knowledge-Based Organizations (IJKBO)* 10 (2): 12-34.

Kross, Ethan, Philippe Verduyn, Gal Sheppes, Cory K Costello, John Jonides, et Oscar Ybarra. 2021. « Social media and well-being: Pitfalls, progress, and next steps ». *Trends in Cognitive Sciences* 25 (1): 55-66.

Krouska, Akrivi, Christos Troussas, et Maria Virvou. 2019. « SN-Learning: An Exploratory Study beyond e-Learning and Evaluation of Its Applications Using EV-SNL Framework ». *Journal of Computer Assisted Learning* 35 (2): 168-77. https://doi.org/10.1111/jcal.12330.

Kulkarni, Akshay, et Adarsha Shivananda. 2021. « Converting Text to Features ». In *Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning Using Python*, édité par Akshay Kulkarni et Adarsha Shivananda, 63-106. Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-7351-7_3.

Kulkarni, Pradnya V, Sunil Rai, et Rohini Kale. 2020. « Recommender system in elearning: a survey ». In , 119-26. Springer.

Kulkarni, Tanay, Madhur Kabra, et Radha Shankarmani. 2019. « User Profiling Based Recommendation System for E-Learning ». In *2019 IEEE 16th India Council International Conference (INDICON)*, 1-4. https://doi.org/10.1109/INDICON47234.2019.9028982.

Kultsova, Marina, Anton Anikin, Irina Zhukova, et Alexander Dvoryankin. 2015. « Ontology-based learning content management system in programming languages domain ». *Communications in Computer and Information Science* 535: 767-77.

Lara, Juan A., David Lizcano, María A. Martínez, Juan Pazos, et Teresa Riera. 2014. « A System for Knowledge Discovery in E-Learning Environments within the European Higher Education Area – Application to Student Data from Open University of Madrid, UDIMA ». *Computers & Education* 72 (mars): 23-36. https://doi.org/10.1016/j.compedu.2013.10.009.

Laurillard, Diana. 2005. « E-learning in higher education ». In *Changing higher education*, 87-100. Routledge.

Lee, Youngju, et Jaeho Choi. 2011. « A review of online course dropout research: Implications for practice and future research ». *Educational Technology Research and Development* 59 (5): 593-618.

Leigh, Doug. 2009. « SWOT Analysis ». In *Handbook of Improving Performance in the Workplace: Volumes 1-3*, 115-40. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470592663.ch24.

Li, Jing, Aixin Sun, Jianglei Han, et Chenliang Li. 2022. « A Survey on Deep Learning for Named Entity Recognition ». *IEEE Transactions on Knowledge and Data Engineering* 34 (1): 50-70. https://doi.org/10.1109/TKDE.2020.2981314.

Li, Wenfeng, Shawkat Ali, Gabriel Lodewijks, Giancarlo Fortino, Giuseppe Di Fatta, Zhouping Yin, Mukaddim Pathan, Antonio Guerrieri, et Qiang Wang, éd. 2016. *Internet and*

Distributed Computing Systems. Vol. 9864. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-45940-0.

Liang, Hong, Xiao Sun, Yunlei Sun, et Yuan Gao. 2017. « Text feature extraction based on deep learning: a review ». *EURASIP Journal on Wireless Communications and Networking* 2017 (1): 211. https://doi.org/10.1186/s13638-017-0993-1.

Liang, Shangsong, Xiangliang Zhang, Zhaochun Ren, et Evangelos Kanoulas. 2018. « Dynamic embeddings for user profiling in twitter ». In , 1764-73.

Lin, Xiaolin, et Rajiv Kishore. 2021. « Social media-enabled healthcare: a conceptual model of social media affordances, online social support, and health behaviors and outcomes ». *Technological Forecasting and Social Change* 166: 120574.

Liu, Sannyuya, Xian Peng, Hercy NH Cheng, Zhi Liu, Jianwen Sun, et Chongyang Yang. 2019. « Unfolding sentimental and behavioral tendencies of learners' concerned topics from course reviews in a MOOC ». *Journal of Educational Computing Research* 57 (3): 670-96.

Liu, Sanya, Cheng Ni, Zhi Liu, Xian Peng, et Hercy NH Cheng. 2017. « Mining individual learning topics in course reviews based on author topic model ». *International Journal of Distance Education Technologies (IJDET)* 15 (3): 1-14.

Liu, Tieyuan, Qiong Wu, Liang Chang, et Tianlong Gu. 2022. « A Review of Deep Learning-Based Recommender System in e-Learning Environments ». *Artificial Intelligence Review* 55 (8): 5953-80. https://doi.org/10.1007/s10462-022-10135-2.

Liu, Yafei, Mingli Zhang, Dan Qi, et Yan Zhang. 2022. « Understanding the role of learner engagement in determining MOOCs satisfaction: A self-determination theory perspective ». *Interactive Learning Environments*, 1-15.

Lu, Jie, Dianshuang Wu, Mingsong Mao, Wei Wang, et Guangquan Zhang. 2015. « Recommender System Application Developments: A Survey ». *Decision Support Systems* 74: 12-32. https://doi.org/10.1016/j.dss.2015.03.008.

Lubis, Fetty Fitriyanti, Yusep Rosmansyah, et Suhono H Supangkat. 2019. « Topic discovery of online course reviews using LDA with leveraging reviews helpfulness ». *International Journal of Electrical and Computer Engineering* 9 (1): 426.

Luna, Vladimir, Rolando Quintero, Miguel Torres, Marco Moreno-Ibarra, Giovanni Guzmán, et Imelda Escamilla. 2015. « An Ontology-Based Approach for Representing the Interaction Process between User Profile and Its Context for Collaborative Learning Environments ». *Computers in Human Behavior* 51 (octobre): 1387-94. https://doi.org/10.1016/j.chb.2014.10.004.

Maedche, Alexander, et Steffen Staab. 2001. « Ontology learning for the semantic web ». *IEEE Intelligent systems* 16 (2): 72-79.

Mak, Myra Yi Ching, Ada Yuen Mei Poon, et Dickson K. W. Chiu. 2022. « Using Social Media as Learning Aids and Preservation: Chinese Martial Arts in Hong Kong ». Chapter. The Digital Folklore of Cyberculture and Digital Humanities. IGI Global. 2022. https://doi.org/10.4018/978-1-6684-4461-0.ch010.

Makhoul, John, Francis Kubala, Richard Schwartz, et Ralph Weischedel. 1999. « Performance measures for information extraction ». In , 249-52. Herndon, VA.

References

Manca, Stefania, et Maria Ranieri. 2016. « Facebook and the Others. Potentials and Obstacles of Social Media for Teaching in Higher Education ». *Computers & Education* 95 (avril): 216-30. https://doi.org/10.1016/j.compedu.2016.01.012.

Mangina, Eleni, et John Kilbride. 2008. « Utilizing Vector Space Models for User Modeling within E-Learning Environments ». *Computers & Education* 51 (2): 493-505. https://doi.org/10.1016/j.compedu.2007.06.008.

Marcot, Bruce G., et Trent D. Penman. 2019. « Advances in Bayesian Network Modelling: Integration of Modelling Technologies ». *Environmental Modelling & Software* 111 (janvier): 386-93. https://doi.org/10.1016/j.envsoft.2018.09.016.

Martin, Florence, Ting Sun, Carl D. Westine, et Albert D. Ritzhaupt. 2022. « Examining Research on the Impact of Distance and Online Learning: A Second-Order Meta-Analysis Study ». *Educational Research Review* 36 (juin): 100438. https://doi.org/10.1016/j.edurev.2022.100438.

McGuinness, Deborah L, et Frank Van Harmelen. 2004. « OWL web ontology language overview ». *W3C recommendation* 10 (10): 2004.

McIntyre, Miranda M, Jessica L Gundlach, et William G Graziano. 2021. « Liking guides learning: The role of interest in memory for STEM topics ». *Learning and Individual Differences* 85: 101960.

Meshi, Dar, Diana I Tamir, et Hauke R Heekeren. 2015. « The emerging neuroscience of social media ». *Trends in cognitive sciences* 19 (12): 771-82.

Michelson, Matthew, et Sofus A Macskassy. 2010. « Discovering users' topics of interest on twitter: a first look ». In , 73-80.

Milićević, Violeta, Nebojša Denić, Zoran Milićević, Ljiljana Arsić, Milica Spasić-Stojković, Dalibor Petković, Jelena Stojanović, Mirjana Krkic, Nataša Sokolov Milovančević, et Aleksandra Jovanović. 2021. « E-learning perspectives in higher education institutions ». *Technological Forecasting and Social Change* 166: 120618.

Missaoui, Siwar, et Ahmed Maalel. 2021. « Student's profile modeling in an adaptive gamified learning environment ». *Education and Information Technologies*, 1-15.

Mittal, Nupur. 2016. « Data, Learning & Privacy in Recommandation Systems », 137.

Moghavvemi, Sedigheh, Tanuosha Paramanathan, Nurliana Md Rahin, et Manal Sharabati. 2017. « Student's perceptions towards using e-learning via Facebook ». *Behaviour & Information Technology* 36 (10): 1081-1100. https://doi.org/10.1080/0144929X.2017.1347201.

Mohamed, Misrah Hamisah, et Michael Hammond. 2018. « MOOCs: a differentiation by pedagogy, content and assessment ». *The International Journal of Information and Learning Technology* 35 (1): 2-11. https://doi.org/10.1108/IJILT-07-2017-0062.

Mohammed, REQQASS, LAKHOUAJA Abdelhak, MAZROUI Azzeddine, et ATIH Idriss. 2015. « AMELIORATION OF THE INTERACTIVE DICTIONARY OF ARABIC LANGUAGE. » *International Journal of Computer Science & Applications* 12 (1).

Moorthy, Krishna, Loh Chun T'ing, Khor Mei Wei, Pauline Tan Zi Mei, Chai Yan Yee, Kevin Lim Jia Wern, et Yue Mei Xin. 2019. « Is Facebook Useful for Learning? A Study in Private Universities in Malaysia ». *Computers & Education* 130 (mars): 94-104. https://doi.org/10.1016/j.compedu.2018.12.002.

Murakami, Riki, et Basabi Chakraborty. 2022. « Investigating the Efficient Use of Word Embedding with Neural-Topic Models for Interpretable Topics from Short Texts ». *Sensors* 22 (3): 852.

Narayan, Vickel, Jan Herrington, et Thom Cochrane. 2019. « Design Principles for Heutagogical Learning: Implementing Student-Determined Learning with Mobile and Social Media Tools ». *Australasian Journal of Educational Technology* 35 (3). https://doi.org/10.14742/ajet.3941.

News, I. B. L. s. d. « MOOCs Skyrocket Due to the Pandemic: Over 180 Million Learners, and Counting | IBL News ». Consulté le 6 décembre 2022. https://iblnews.org/moocs-skyrocket-due-to-the-pandemic-over-180-million-learners-and-counting/.

Nguyen, Loc, et Phung Do. 2009. « Combination of Bayesian Network and Overlay Model in User Modeling ». In *Computational Science – ICCS 2009*, édité par Gabrielle Allen, Jarosław Nabrzyski, Edward Seidel, Geert Dick van Albada, Jack Dongarra, et Peter M. A. Sloot, 5-14. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-01973-9_2.

Nilashi, Mehrbakhsh, Othman Ibrahim, et Karamollah Bagherifard. 2018. « A Recommender System Based on Collaborative Filtering Using Ontology and Dimensionality Reduction Techniques ». *Expert Systems with Applications* 92 (février): 507-20. https://doi.org/10.1016/j.eswa.2017.09.058.

Njenga, James Kariuki, et Louis Cyril Henry Fourie. 2010. « The myths about e-learning in higher education ». *British journal of educational technology* 41 (2): 199-212.

Onan, Aytuğ, et Mansur Alp Toçoğlu. 2021. « Weighted word embeddings and clustering-based identification of question topics in MOOC discussion forum posts ». *Computer Applications in Engineering Education* 29 (4): 675-89.

Orlandi, Fabrizio, John Breslin, et Alexandre Passant. 2012. « Aggregated, interoperable and multi-domain user profiles for the social web ». In *Proceedings of the 8th International Conference on Semantic Systems*, 41-48. I-SEMANTICS '12. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2362499.2362506.

Otter, Daniel W., Julian R. Medina, et Jugal K. Kalita. 2021. « A Survey of the Usages of Deep Learning for Natural Language Processing ». *IEEE Transactions on Neural Networks and Learning Systems* 32 (2): 604-24. https://doi.org/10.1109/TNNLS.2020.2979670.

Ouaftouh, Sara, Ahmed Zellou, et Ali Idri. 2015. « User profile model: A user dimension based classification ». In *2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA)*, 1-5. https://doi.org/10.1109/SITA.2015.7358378.

Ouatiq, Amina, Kamal El Guemmat, Khalifa Mansouri, et Mohammed Qbadou. 2021. « Towards an ontological learner's modeling during and after the COVID-19 pandemic ». *International Journal of Advanced Computer Science and Applications* 12 (2).

Pahl-Wostl, Claudia, et Matt Hare. 2004. « Processes of Social Learning in Integrated Resources Management ». *Journal of Community & Applied Social Psychology* 14 (3): 193-206. https://doi.org/10.1002/casp.774.

Panagiotakopoulos, Theodor, Sotiris Kotsiantis, Georgios Kostopoulos, Omiros Iatrellis, et Achilles Kameas. 2021. « Early Dropout Prediction in MOOCs through Supervised

Learning and Hyperparameter Optimization ». *Electronics* 10 (14): 1701. https://doi.org/10.3390/electronics10141701.

Park, Jinsoo, Kimoon Sung, et Sewon Moon. 2008. « Developing Graduation Screen Ontology Based on the METHONTOLOGY Approach ». In *2008 Fourth International Conference on Networked Computing and Advanced Information Management*, 2:375-80. https://doi.org/10.1109/NCM.2008.215.

Pathak, Ajeet Ram, Manjusha Pandey, et Siddharth Rautaray. 2021. « Topic-level sentiment analysis of social media data using deep learning ». *Applied Soft Computing* 108: 107440.

Peng, Xian, Chengyang Han, Fan Ouyang, et Zhi Liu. 2020. « Topic tracking model for analyzing student-generated posts in SPOC discussion forums ». *International Journal of Educational Technology in Higher Education* 17 (1): 1-22.

Peng, Xian, Sanya Liu, Zhi Liu, Wenbin Gan, et Jianwen Sun. 2016. « Mining learners' topic interests in course reviews based on like-LDA model ». *International Journal of Innovative Computing, Information and Control* 12 (6): 2099-2110.

Peter Ozioma, U, A Bethran Chibuike, A Alphonsus Onyekachi, et A VC. 2022. « Development of a Visual Semantic Web Ontology Based Learning Management System ». *International Journal of Engineering Applied Sciences and Technology* 6 (10): 226-38.

Piao, Guangyuan, et John G Breslin. 2017. « Inferring user interests for passive users on twitter by leveraging followee biographies ». In , 122-33. Springer.

Pireva, Krenare, et Petros Kefalas. 2018. « A Recommender System Based on Hierarchical Clustering for Cloud E-Learning ». In *Intelligent Distributed Computing XI*, édité par Mirjana Ivanović, Costin Bădică, Jürgen Dix, Zoran Jovanović, Michele Malgeri, et Miloš Savić, 235-45. Studies in Computational Intelligence. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-66379-1_21.

Popp, Heribert, et Dieter Lödel. 1995. « Fuzzy Techniques and User Modeling in Sales Assistants ». *User Modeling and User-Adapted Interaction* 5 (3): 349-70. https://doi.org/10.1007/BF01126115.

« protégé ». s. d. Consulté le 22 juillet 2018. https://protege.stanford.edu/.

Qi, Qianqian, David J Hessen, et Peter GM van der Heijden. 2021. « A Comparison of Latent Semantic Analysis and Correspondence Analysis for Text Mining ». *arXiv preprint arXiv:2108.06197*.

Ramírez-Gallego, Sergio, Bartosz Krawczyk, Salvador García, Michał Woźniak, et Francisco Herrera. 2017. « A Survey on Data Preprocessing for Data Stream Mining: Current Status and Future Directions ». *Neurocomputing* 239 (mai): 39-57. https://doi.org/10.1016/j.neucom.2017.01.078.

Raulji, Jaideepsinh K, et Jatinderkumar R Saini. 2016. « Stop-word removal algorithm and its implementation for Sanskrit language ». *International Journal of Computer Applications* 150 (2): 15-17.

Ravenscroft, James, Arie Cattan, Amanda Clare, Ido Dagan, et Maria Liakata. 2021. « CD2CR: Co-reference Resolution Across Documents and Domains ». arXiv. https://doi.org/10.48550/arXiv.2101.12637.

Rawashdeh, Alaa Zuhir Al, Enaam Youssef Mohammed, Asma Rebhi Al Arab, Mahmoud Alara, Butheyna Al-Rawashdeh, et Butheyna Al-Rawashdeh. 2021. « Advantages and

Disadvantages of Using E-Learning in University Education: Analyzing Students' Perspectives ». *Electronic Journal of E-Learning* 19 (3): 107-17. https://doi.org/10.34190/ejel.19.3.2168.

Reimers, Nils, et Iryna Gurevych. 2019. « Sentence-bert: Sentence embeddings using siamese bert-networks ». *arXiv preprint arXiv:1908.10084*.

Rezgui, Kalthoum, Hédia Mhiri, et Khaled Ghédira. 2014. « An Ontology-based Profile for Learner Representation in Learning Networks ». *International Journal of Emerging Technologies in Learning (iJET)* 9 (3): 16. https://doi.org/10.3991/ijet.v9i3.3305.

Ribeiro, Alexandre M., João LT Silva, Elisa Boff, Tiago Primo, et Rosa M. Viccari. 2011. « Towards ontological profiles in communities of practice ». In *Proceedings of 4th Brazilian Workshop on Semantic WebandEducation (WSWEd'2011) at 22nd Brazilian Symposium in Informatics in Education (SBIE'2011)*, 2334-43.

Rich, Elaine. 1979. « User Modeling via Stereotypes ». *Cognitive Science* 3 (4): 329-54. https://doi.org/10.1016/S0364-0213(79)80012-9.

Rim, Rebai, Maalej Mohamed Amin, et Mahfoudhi Adel. 2013. « Bayesian networks for user modeling: Predicting the user's preferences ». In *13th International Conference on Hybrid Intelligent Systems (HIS 2013)*, 144-48. https://doi.org/10.1109/HIS.2013.6920472.

Robin, Rene, et G. Uma. 2010. « A Novel Algorithm for Fully Automated Ontology Merging Using Hybrid Strategy ». *European Journal of Scientific Research* 47 (novembre): 074-081.

Rodrigues, Helena, Filomena Almeida, Vanessa Figueiredo, et Sara L Lopes. 2019. « Tracking e-learning through published papers: A systematic review ». *Computers & Education* 136: 87-98.

Rosé, Carolyn Penstein, Ryan Carlson, Diyi Yang, Miaomiao Wen, Lauren Resnick, Pam Goldman, et Jennifer Sherer. 2014. « Social factors that contribute to attrition in MOOCs ». In , 197-98.

Rosenberg, Marc J., et Rob Foshay. 2002. « E-Learning: Strategies for Delivering Knowledge in the Digital Age ». *Performance Improvement* 41 (5): 50-51. https://doi.org/10.1002/pfi.4140410512.

Rosner, Frank, Alexander Hinneburg, Michael Röder, Martin Nettling, et Andreas Both. 2014. « Evaluating topic coherence measures ». *arXiv preprint arXiv:1403.6397*.

Rumjaun, Anwar, et Fawzia Narod. 2020. « Social Learning Theory—Albert Bandura ». In *Science Education in Theory and Practice: An Introductory Guide to Learning Theory*, édité par Ben Akpan et Teresa J. Kennedy, 85-99. Springer Texts in Education. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-43620-9_7.

Saadatdoost, Robab, Alex Tze Hiang Sim, Hosein Jafarkarimi, et Jee Mei Hee. 2015. « Exploring MOOC from education and Information Systems perspectives: a short literature review ». *Educational Review* 67 (4): 505-18. https://doi.org/10.1080/00131911.2015.1058748.

Sadri, Arif Mohaimin, Samiul Hasan, et Satish V Ukkusuri. 2019. « Joint inference of user community and interest patterns in social interaction networks ». *Social Network Analysis and Mining* 9: 1-23.

Salloum, Said A., Nafla Mahdi Nasser AlAhbabi, Mohammed Habes, Ahmad Aburayya, et Iman Akour. 2021. « Predicting the Intention to Use Social Media Sites: A Hybrid SEM - Machine Learning Approach ». In *Advanced Machine Learning Technologies and Applications*, édité par Aboul-Ella Hassanien, Kuo-Chi Chang, et Tang Mincong, 324-34. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-69717-4_32.

Salloum, Said A., Mostafa Al-Emran, Mohammed Habes, Mahmoud Alghizzawi, Mazuri Abd. Ghani, et Khaled Shaalan. 2021. « What Impacts the Acceptance of E-Learning Through Social Media? An Empirical Study ». In *Recent Advances in Technology Acceptance Models and Theories*, édité par Mostafa Al-Emran et Khaled Shaalan, 419-31. Studies in Systems, Decision and Control. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-64987-6_24.

Salmerón, Ladislao, Arantxa García, et Eduardo Vidal-Abarca. 2018. « The development of adolescents' comprehension-based Internet reading activities ». *Learning and Individual Differences* 61: 31-39.

Salton, Gerard, Anita Wong, et Chung-Shu Yang. 1975. « A vector space model for automatic indexing ». *Communications of the ACM* 18 (11): 613-20.

Sambrook, Sally. 2003. « E-learning and small organisations ». In *Human Resource Development in Small Organisations*, 199-228. Routledge.

Sangaraju, Vasudeva Raju, Bharath Kumar Bolla, Deepak Kumar Nayak, et Jyothsna Kh. 2022. « Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings ». arXiv. https://doi.org/10.48550/arXiv.2205.07259.

Sangrà, Albert, Dimitrios Vlachopoulos, et Nati Cabrera. 2012. « Building an inclusive definition of e-learning: An approach to the conceptual framework ». *International Review of Research in Open and Distributed Learning* 13 (2): 145-59.

Sarwar, Badrul, George Karypis, Joseph Konstan, et John Reidl. 2001. « Item-Based Collaborative Filtering Recommendation Algorithms ». In , 285-95. ACM Press. https://doi.org/10.1145/371920.372071.

Sarwari, Kawita, Ahmad Fawad Kakar, Jawad Golzar, et Mir Abdullah Miri. 2022. « Distance learning during COVID-19 in Afghanistan: Challenges and opportunities ». *E-Learning and Digital Media* 19 (2): 144-62. https://doi.org/10.1177/20427530211044757.

Sateli, Bahar, Felicitas Löffler, Birgitta König-Ries, et René Witte. 2017. « ScholarLens: Extracting Competences from Research Publications for the Automatic Generation of Semantic User Profiles ». *PeerJ Computer Science* 3 (juillet): e121. https://doi.org/10.7717/peerj-cs.121.

Schneider, Samantha L., et Martha Laurin Council. 2021. « Distance Learning in the Era of COVID-19 ». *Archives of Dermatological Research* 313 (5): 389-90. https://doi.org/10.1007/s00403-020-02088-9.

Schöch, Christof. 2021. « Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama ». arXiv. https://doi.org/10.48550/arXiv.2103.13019.

Shapiro, Heather B, Clara H Lee, Noelle E Wyman Roth, Kun Li, Mine Çetinkaya-Rundel, et Dorian A Canelas. 2017. « Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers ». *Computers & Education* 110: 35-50.

Sharma, Anuj, Nripendra P Rana, et Robin Nunkoo. 2021. « Fifty years of information management research: A conceptual structure analysis using structural topic modeling ». *International Journal of Information Management* 58: 102316.

Shawky, Doaa, et Ashraf Badawi. 2019. « Towards a personalized learning experience using reinforcement learning ». In *Machine learning paradigms: Theory and application*, 169-87. Springer.

Sheeba, T., et S. Hameetha Begum. 2013. « An Ontology Approach To Represent User Profiles in E-Learning ». *International Journal of Engineering Research & Technology (IJERT)*.

Sheeba, T., et Reshmy Krishnan. 2014. « Ontology Based Semantic Retrieval of Learner Profile ». In *The International Arab Conference on Information Technology.(ACIT2014)*.

Sheeba, T, et Reshmy Krishnan. 2015. « Semantic retrieval based on SPARQL and SWRL for learner profile ». *Int J Appl Eng Res* 10: 34549-54.

Sheeba, T., et Reshmy Krishnan. 2016. « An Ontological Framework of Semantic Learner Profile in an E-Learning System ». In *Advances in Brain Inspired Cognitive Systems*, édité par Cheng-Lin Liu, Amir Hussain, Bin Luo, Kay Chen Tan, Yi Zeng, et Zhaoxiang Zhang, 10023:284-97. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-49685-6_26.

Sheeba, T., & Krishnan, R.2020. « A Semantic Approach of Building Dynamic Learner Profile Model Using WordNet ». In *Advanced Computing and Intelligent Engineering*, édité par Bibudhendu Pati, Chhabi Rani Panigrahi, Rajkumar Buyya, et Kuan-Ching Li, 263-72. Advances in Intelligent Systems and Computing. Singapore: Springer. https://doi.org/10.1007/978-981-15-1081-6_22.

Shishehchi, Saman, Seyed Yashar Banihashem, et Nor Azan Mat Zin. 2010. « A proposed semantic recommendation system for e-learning: A rule and ontology based e-learning recommendation system ». In *2010 International Symposium on Information Technology*, 1:1-5. https://doi.org/10.1109/ITSIM.2010.5561329.

Sidorov, Grigori, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, et Liliana Chanona-Hernández. 2014. « Syntactic N-Grams as Machine Learning Features for Natural Language Processing ». *Expert Systems with Applications*, Methods and Applications of Artificial and Computational Intelligence, 41 (3): 853-60. https://doi.org/10.1016/j.eswa.2013.08.015.

« SIOC Core Ontology Specification ». s. d. Consulté le 6 janvier 2020. https://www.w3.org/Submission/sioc-spec/.

Siva Rama Rao, Akula VS, PVV Vamsi, N Rashmika, K Hemanth, et K Aditya Kumar. 2022. « Named Entity Recognition Using Stanford Classes and NLTK ». In , 583-97. Springer.

Skillen, Kerry-Louise, Liming Chen, Chris D. Nugent, Mark P. Donnelly, William Burns, et Ivar Solheim. 2014. « Ontological User Modelling and Semantic Rule-Based Reasoning for Personalisation of Help-On-Demand Services in Pervasive Environments ». *Future Generation Computer Systems* 34 (mai): 97-109. https://doi.org/10.1016/j.future.2013.10.027.

Slimani, Thabet. 2015. « Ontology development: A comparing study on tools, languages and formalisms ». *Indian Journal of Science and Technology* 8 (24).

References

Smith, Becky, et Min Eng. 2013. « MOOCs: A Learning Journey ». In *Hybrid Learning and Continuing Education*, édité par Simon K. S. Cheung, Joseph Fong, Wilfred Fong, Fu Lee Wang, et Lam For Kwok, 244-55. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39750-9_23.

Song, Xuemeng, Liqiang Nie, Luming Zhang, Mohammad Akbari, et Tat-Seng Chua. 2015. « Multiple social network learning and its application in volunteerism tendency prediction ». In , 213-22.

Staab, Steffen, et Rudi Studer, éd. 2009. *Handbook on Ontologies*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-92673-3.

Suardika, I. Ketut, Alberth, Mursalim, Siam, Lelly Suhartini, et Nikolaus Pasassung. 2020. « Using WhatsApp for Teaching a Course on the Education Profession: Presence, Community and Learning ». *International Journal of Mobile and Blended Learning (IJMBL)* 12 (1): 17-32. https://doi.org/10.4018/IJMBL.2020010102.

Suleman, Raja Muhammad, et Ioannis Korkontzelos. 2021. « Extending latent semantic analysis to manage its syntactic blindness ». *Expert Systems with Applications* 165: 114130.

Suzianti, Amalia, et Sabrina Ayu Paramadini. 2021. « Continuance Intention of E-Learning: The Condition and Its Connection with Open Innovation ». *Journal of Open Innovation: Technology, Market, and Complexity* 7 (1): 97. https://doi.org/10.3390/joitmc7010097.

Symeonidis, P., et D. Malakoudis. 2016. « MoocRec.com : Massive open online courses recommender system ». In . Vol. 1688. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84991087056&partnerID=40&md5=2b23a08488660546a6065d0c381f7ada.

Talaghzi, Jallal, Mostafa Bellafkih, Abdellah Bennane, Mohammed Majid Himmi, et Mohammed Amraouy. 2023. « A Combined E-Learning Course Recommender System ». *International Journal of Emerging Technologies in Learning (Online)* 18 (6): 53.

Tao, Da, Pei Fu, Yunhui Wang, Tingru Zhang, et Xingda Qu. 2022. « Key characteristics in designing massive open online courses (MOOCs) for user acceptance: an application of the extended technology acceptance model ». *Interactive Learning Environments* 30 (5): 882-95. https://doi.org/10.1080/10494820.2019.1695214.

Tao, Ke, Fabian Abel, Qi Gao, et Geert-Jan Houben. 2011. « TUMS: Twitter-Based User Modeling Service. » In , 7117:269-83. Springer.

Tarus, John K., Zhendong Niu, et Dorothy Kalui. 2018. « A Hybrid Recommender System for E-Learning Based on Context Awareness and Sequential Pattern Mining ». *Soft Computing* 22 (8): 2449-61. https://doi.org/10.1007/s00500-017-2720-6.

Tarus, John K., Zhendong Niu, et Ghulam Mustafa. 2018a. « Knowledge-Based Recommendation: A Review of Ontology-Based Recommender Systems for e-Learning ». *Artificial Intelligence Review* 50 (1): 21-48. https://doi.org/10.1007/s10462-017-9539-5.

Tarus, J. K., Niu, Z., & Mustafa, G.2018b. « Knowledge-Based Recommendation: A Review of Ontology-Based Recommender Systems for e-Learning ». *Artificial Intelligence Review* 50 (1): 21-48. https://doi.org/10.1007/s10462-017-9539-5.

Tarus, John K., Zhendong Niu, et Abdallah Yousif. 2017. « A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining ». *Future Generation Computer Systems* 72 (juillet): 37-48. https://doi.org/10.1016/j.future.2017.02.049.

Tasharofi, Samira, Fahimeh Raja, Farhad Oroumchian, et Masoud Rahgozar. 2007. « Evaluation of statistical part of speech tagging of persian text ». In *2007 9th International Symposium on Signal Processing and Its Applications*, 1-4. https://doi.org/10.1109/ISSPA.2007.4555312.

Teoli, Dac, Terrence Sanvictores, et Jason An. 2022. « SWOT Analysi ». https://www.ncbi.nlm.nih.gov/books/NBK537302.

Terragni, Silvia, Elisabetta Fersini, Bruno Giovanni Galuzzi, Pietro Tropeano, et Antonio Candelieri. 2021. « OCTIS: Comparing and Optimizing Topic models is Simple! » In , 263-70.

Terzi, Banu, Seçil Bulut, et Nurten Kaya. 2019. « Factors affecting nursing and midwifery students' attitudes toward social media ». *Nurse education in practice* 35: 141-49.

Thavareesan, Sajeetha, et Sinnathamby Mahesan. 2020. « Word embedding-based Part of Speech tagging in Tamil texts ». In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, 478-82. https://doi.org/10.1109/ICIIS51140.2020.9342640.

« The Organization Ontology ». s. d. Consulté le 22 juillet 2023. https://www.w3.org/TR/vocab-org/.

Tîrziu, Andreea-Maria, et Cătălin Vrabie. 2015. « Education 2.0: E-Learning Methods ». *Procedia - Social and Behavioral Sciences*, The Proceedings of 5th World Conference on Learning, Teaching and Educational Leadership, 186 (mai): 376-80. https://doi.org/10.1016/j.sbspro.2015.04.213.

Troussas, Christos, Akrivi Krouska, et Cleo Sgouropoulou. 2021. « Impact of social networking for advancing learners' knowledge in E-learning environments ». *Education and Information Technologies* 26 (4): 4285-4305.

Upadhyaya, K T, et Debasis Mallik. 2013. « E-Learning as a Socio-Technical System: An Insight into Factors Influencing Its Effectiveness ». *Business Perspectives and Research* 2 (1): 1-12. https://doi.org/10.1177/2278533720130101.

Vayansky, Ike, et Sathish A. P. Kumar. 2020. « A Review of Topic Modeling Methods ». *Information Systems* 94 (décembre): 101582. https://doi.org/10.1016/j.is.2020.101582.

Voutilainen, Atro. 2003. *Part-of-speech tagging*. Vol. 219. The Oxford handbook of computational linguistics.

Vuong, Quan-Hoang, Viet-Phuong La, Minh-Hoang Nguyen, Manh-Toan Ho, Manh-Tung Ho, et Peter Mantello. 2020. « Improving Bayesian Statistics Understanding in the Age of Big Data with the Bayesvl R Package ». *Software Impacts* 4 (mai): 100016. https://doi.org/10.1016/j.simpa.2020.100016.

Wang, Wei, Lihuan Guo, Ling He, et Yenchun Jim Wu. 2019. « Effects of Social-Interactive Engagement on the Dropout Ratio in Online Learning: Insights from MOOC ». *Behaviour & Information Technology* 38 (6): 621-36. https://doi.org/10.1080/0144929X.2018.1549595.

Wang, Yinglin, Ming Wang, et Hamido Fujita. 2020. « Word Sense Disambiguation: A Comprehensive Knowledge Exploitation Framework ». *Knowledge-Based Systems* 190 (février): 105030. https://doi.org/10.1016/j.knosys.2019.105030.

Waqialla, Mohammad, et Muhammad Imran Razzak. 2016. « An Ontology-based Framework Aiming to Support Cardiac Rehabilitation Program. » In , 23-32.

Webb, Geoffrey I., Michael J. Pazzani, et Daniel Billsus. 2001. « Machine Learning for User Modeling ». *User Modeling and User-Adapted Interaction* 11 (1): 19-29. https://doi.org/10.1023/A:1011117102175.

Wentling, Tim L, Consuelo Waight, James Gallaher, Jason La Fleur, Christine Wang, et Alaina Kanfer. 2000. « E-learning: A review of literature ». *Knowledge and Learning Systems Group NCSA* 9 (1): 73.

Wong, Jacqueline, Martine Baars, Dan Davis, Tim Van Der Zee, Geert-Jan Houben, et Fred Paas. 2019. « Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review ». *International Journal of Human–Computer Interaction* 35 (4-5): 356-73. https://doi.org/10.1080/10447318.2018.1543084.

Xiong, Yao, Hongli Li, Mindy L Kornhaber, Hoi K Suen, Barton Pursel, et Deborah D Goins. 2015. « Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. » *Global Education Review* 2 (3): 23-33.

Xu, Zhiheng, Long Ru, Liang Xiang, et Qing Yang. 2011. « Discovering user interest on twitter with a modified author-topic model ». In , 1:422-29. IEEE.

Yaakub, Mohd Ridzwan, Muhammad Iqbal Abu Latiffi, et Liyana Safra Zaabar. 2019. « A Review on Sentiment Analysis Techniques and Applications ». *IOP Conference Series: Materials Science and Engineering* 551 (1): 012070. https://doi.org/10.1088/1757-899X/551/1/012070.

Yadav, Vikas, et Steven Bethard. 2019. « A Survey on Recent Advances in Named Entity Recognition from Deep Learning models ». arXiv. https://doi.org/10.48550/arXiv.1910.11470.

Yago, Hector, Julia Clemente, Daniel Rodriguez, et Pedro Fernandez-de-Cordoba. 2018. « ON-SMMILE: Ontology Network-Based Student Model for MultIple Learning Environments ». *Data & Knowledge Engineering* 115 (mai): 48-67. https://doi.org/10.1016/j.datak.2018.02.002.

Yang, Bokai, Hengtao Tang, Ling Hao, et John R Rose. 2022. « Untangling chaos in discussion forums: A temporal analysis of topic-relevant forum posts in MOOCs ». *Computers & Education* 178: 104402.

Yang, Diyi, Miaomiao Wen, et Carolyn Rose. 2014. « Peer influence on attrition in massively open online courses ». In .

Yang, Xi, Kaiwen Yang, Tianxu Cui, Min Chen, et Liyan He. 2022. « A Study of Text Vectorization Method Combining Topic Model and Transfer Learning ». *Processes* 10 (2): 350.

Yang, Yuqin, Qiyun Wang, Huay Lit Woo, et Choon Lang Quek. 2011. « Using Facebook for teaching and learning: a review of the literature ». *International Journal of Continuing Engineering Education and Life Long Learning* 21 (1): 72-86. https://doi.org/10.1504/IJCEELL.2011.039695.

Yılmaz, Fatma Gizem Karaoğlan, et Ramazan Yılmaz. 2022. « Exploring the role of sociability, sense of community and course satisfaction on students' engagement in flipped classroom supported by facebook groups ». *Journal of Computers in Education*, 1-28.

Yu, Youqin, et Bicheng Li. 2021. « Microblog User Interest Recognition Based on Multi-Granularity Text Feature Representation ». In , 1-10.

Yulianti, Lenny Putri, et Kridanto Surendro. 2018. « Ontology Model for Indigenous Knowledge ». In *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, 231-36. IEEE.

Zankadi, Hajar, Imane Hilal, Najima Daoudi, et Abdellah Idrissi. 2018. « Facebook and moocs: a comparative analysis for a collaborative learning ». In , 1-7. IEEE.

Zankadi, Hajar, Imane Hilal, Abdellah Idrissi, et Najima Daoudi. 2022. « A Social Profile Ontology to Enhance Learner Experience in MOOCs. » *International Journal of Emerging Technologies in Learning* 17 (4).

Zankadi, Hajar, Abdellah Idrissi, Najima Daoudi, et Imane Hilal. 2022. « Identifying learners' topical interests from social media content to enrich their course preferences in MOOCs using topic modeling and NLP techniques ». *Education and Information Technologies*, 1-18.

Zapata-Rivera, Diego, et Burcu Arslan. 2021. « Enhancing Personalization by Integrating Top-Down and Bottom-Up Approaches to Learner Modeling ». In *Adaptive Instructional Systems. Adaptation Strategies and Methods*, édité par Robert A. Sottilare et Jessica Schwarz, 234-46. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-77873-6_17.

Zaware, Sarika, Deep Patadiya, Abhishek Gaikwad, Sanket Gulhane, et Akash Thakare. 2021. « Text summarization using tf-idf and textrank algorithm ». In , 1399-1407. IEEE.

Zhang, Aimin, Yingjun Zhang, et Junzhi Jia. 2023. « The reuse of DCMI metadata terms in linked open vocabulary ». *The Electronic Library* 41 (1): 152-65.

Zhang, Fu, Ke Wang, Zhiyin Li, et Jingwei Cheng. 2019. « Temporal Data Representation and Querying Based on RDF ». *IEEE Access* 7: 85000-23. https://doi.org/10.1109/ACCESS.2019.2924550.

Zhang, Junsong, Wu Zhao, Gang Xie, et Hong Chen. 2011. « Ontology- Based Knowledge Management System and Application ». *Procedia Engineering* 15: 1021-29. https://doi.org/10.1016/j.proeng.2011.08.189.

Zhang, Qing, Kyle L Peck, Adelina Hristova, Kathryn W Jablokow, Vicki Hoffman, Eunsung Park, et Rebecca Yvonne Bayeck. 2016. « Exploring the communication preferences of MOOC learners and the value of preference-based groups: Is grouping enough? » *Educational Technology Research and Development* 64 (4): 809-37.

Zhao, Rui, et Kezhi Mao. 2018. « Fuzzy Bag-of-Words Model for Document Representation ». *IEEE Transactions on Fuzzy Systems* 26 (2): 794-804. https://doi.org/10.1109/TFUZZ.2017.2690222.

Zheng, Saijing, Mary Beth Rosson, Patrick C. Shih, et John M. Carroll. 2015. « Understanding Student Motivation, Behaviors and Perceptions in MOOCs ». In , 1882-95. ACM Press. https://doi.org/10.1145/2675133.2675217.

Zhou, Lina. 2007. « Ontology Learning: State of the Art and Open Issues ». *Information Technology and Management* 8 (3): 241-52. https://doi.org/10.1007/s10799-007-0019-5.

Zhou, Yizhuo, Jin Zhao, et Jianjun Zhang. 2020. « Prediction of learners' dropout in E-learning based on the unusual behaviors ». *Interactive Learning Environments* 0 (0): 1-25. https://doi.org/10.1080/10494820.2020.1857788.

Zine, Oumaima, Abdelouafi Meziane, et Mohamed Boudchiche. 2017. « Towards a high-quality lemma-based text to speech system for the arabic language ». In , 53-66. Springer.

Zukerman, Ingrid, et David W. Albrecht. 2001. « Predictive Statistical Models for User Modeling ». *User Modeling and User-Adapted Interaction* 11 (1): 5-18. https://doi.org/10.1023/A:1011175525451.

Zulkanain, Nani Amalina, Suraya Miskon, et Norris Syed Abdullah. 2020. « An Adapted Pedagogical Framework in Utilizing WhatsApp for Learning Purpose ». *Education and Information Technologies* 25 (4): 2811-22. https://doi.org/10.1007/s10639-019-10096-0.

Zuo, Xinyu, Yubo Chen, Kang Liu, et Jun Zhao. 2019. « Event Co-Reference Resolution via a Multi-Loss Neural Network without Using Argument Information ». *Science China Information Sciences* 62 (11): 212101. https://doi.org/10.1007/s11432-018-9833-1.