

Modélisation conceptuelle à l'ère de la digitalisation : Avancées de la recherche en France

Conceptual Modeling in the Era of Digitalization: Research Advances in France

Ramona Elali¹, Solweig Hennechart², Chloé Vigneau³, Elena Kornysheva⁴

¹ CRI, Université Paris 1 Panthéon - Sorbonne, France, ramona.elali@univ-paris1.fr

² LRSV, IRIT, Université Toulouse 3 Paul Sabatier, solweig.hennechart@univ-tlse3.fr

³ LLR, Ecole Polytechnique, CEDRIC, CNAM, France, chloe.vigneau@polytechnique.edu

⁴ CEDRIC, CNAM, France, elena.kornysheva@cnam.fr

RÉSUMÉ. Cet article représente un recueil de trois meilleurs articles du Forum des Jeunes Chercheuses Jeunes Chercheurs (JCJC) de la conférence INFORSID 2022. Ces contributions comportent la modélisation qui a été sélectionnée comme axe central de cet article invité. Les travaux des JCJC français portent sur l'utilisation de la modélisation conceptuelle pour des sujets variés tels que la construction des modèles de processus contextuels, la modélisation de chimio-diversité et les jeux d'éducation en classe. Dans le cadre de cet article, les JCJC expliquent le contexte de leur recherche, présentent le modèle conceptuel correspondant, expliquent l'utilisation du modèle et les perspectives de recherche.

ABSTRACT. This paper represents a collection of three best articles from the Forum des Jeunes Chercheuses Jeunes Chercheurs (JCJC) – Young Researchers – of the INFORSID 2022 conference. These contributions include the modeling that has been selected as the central topic of this invited paper. The works of the French Young Researchers focus on the use of conceptual modeling for various topics such as construction of contextual process models, modeling of chemodiversity, and educational games in classrooms. In this article, the Young Researchers explain the context of their research, present the corresponding conceptual model, explain the use of the model and the research perspectives.

MOTS-CLÉS. Modélisation conceptuelle, modèle de processus contextuel, modèle de chimio-diversité, modèle de jeu en classe.

KEYWORDS. Conceptual modelling, contextual process model, chemical diversity model, classroom game model.

1. Introduction

La modélisation conceptuelle vise la construction des représentations formelles d'un domaine [WAN 02]. Pour construire un modèle conceptuel il faut abstraire du monde réel des informations pertinentes par rapport à un objectif donné. La modélisation conceptuelle est utilisée dans plusieurs domaines : conception des bases de données, développement des logiciels, gestion des processus métiers, etc. [FET 209]. Ce domaine existe depuis des décennies, cependant, avec le développement des technologies actuelles, elle demeure d'actualité et instancie même de nouveaux défis associés aux données massives, à l'intelligence artificielle, à l'apprentissage automatique, etc. [REC 21] [HAR 20]. Les domaines d'application de la modélisation conceptuelle se démultiplient et de nouvelles façons d'utiliser les modèles apparaissent.

Cet article vise à présenter les travaux récents liés à la modélisation conceptuelle en France. Il inclut les contributions des doctorants qui ont participé au Forum des Jeunes Chercheuses Jeunes Chercheurs (JCJC) de la conférence INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision) qui a eu lieu en Mai-Juin 2022 à Dijon (<https://inforsid2022.sciencesconf.org/>). Le Forum s'est déroulé sous forme de « Dragons/Chevaliers » où chaque doctorant a exposé ses travaux, tout en jouant un rôle de dragon en attaquant une présentation et un rôle de chevalier en protégeant une autre. Cela a permis d'avoir une discussion très riche autour des travaux des doctorants, mais également de leur donner la possibilité de formuler et d'exprimer un regard critique sur les travaux de leurs collègues. Sur douze travaux du Forum JCJC, trois meilleures contributions (toutes des Jeunes

Chercheuses) ont été sélectionnés pour faire partie de cet article invité. Ces trois contributions incluent la modélisation conceptuelle en base commune structurante. Chaque doctorante a retravaillé son article pour présenter le contexte de sa recherche, le modèle développé correspondant, expliquer l'utilisation de son modèle et donner ses perspectives de recherche.

La première contribution (Section 2) porte sur l'utilisation des modèles conceptuels pour les recommandations contextuelles basées sur la fouille des intentions et des processus [ELA 22]. Ramona Elali présente un modèle de processus contextuel, obtenu à partir des logs d'activités en utilisant les techniques de fouille de processus et enrichi par les informations sur le contexte issues des capteurs environnementaux dans le cas d'une maison intelligente (Smart Home).

Les travaux de recherche de Solweig Hennechart visent à représenter la chimio-diversité du vivant en modélisant les informations sur les composés d'origines biologiques [HEN 22]. Elle conçoit une base de données pour inclure ces informations et les métadonnées associées. Plusieurs sources de données hétérogènes doivent être incluses et complétées en utilisant les modèles prédictifs. Cela aboutira à terme à la modélisation de l'origine biologique et à la classification chimique des composés du vivant (Section 3).

La dernière contribution concerne la création de jeu vidéo par des élèves en classe de seconde au lycée [VIG 22]. Chloé Vigneau élabore un modèle qui permet de suivre et d'évaluer les activités des élèves, en mettant en évidence leur progression et les compétences acquises. Le système est fondé sur les tâches de production et d'apprentissage réalisées par les élèves (Section 4).

Les sections suivantes présentent les trois travaux de recherche sélectionnés. Nous concluons notre article dans la Section 5.

2. Recommandation contextuelles basées sur la fouille d'intentions et les ontologies

L'objectif principal des entreprises est d'améliorer leurs processus métiers pour augmenter leurs revenus. La fouille d'intentions permet de fournir de nouveaux services avec une meilleure qualité car elle se concentre sur la partie intentionnelle d'un processus, plus proche du raisonnement de l'utilisateur. Les principaux objectifs de la fouille d'intentions sont d'analyser le comportement des utilisateurs d'une manière fiable et de fournir des recommandations de bonne qualité. Généralement, l'activité de l'utilisateur est affectée par plusieurs facteurs contextuels. Ces informations contextuelles peuvent être obtenues à partir de nombreuses sources de données externes. Ce travail explore comment il est possible de combiner différents types de sources riches en données de contexte afin d'obtenir une meilleure qualité des recommandations. Bien que les recherches existantes se concentrent principalement sur des données d'activité standard, quelques-unes utilisent les ontologies pour combiner les diverses sources de données. Nous proposons une nouvelle approche qui combine plusieurs types de sources dans des ontologies de domaine, utilise ces ontologies pour construire le modèle de processus intentionnel, et ensuite se sert du modèle intentionnel pour faire des recommandations contextuelles [ELA 22].

2.1. Contexte

Aujourd'hui, avec l'évolution des technologies de l'information et la présence de l'Internet les données sont devenues disponibles partout. Cependant, la plupart de ces données ne sont ni structurées ni bien organisées [SIE 14]. Par conséquent, le problème principal revient à identifier ce que l'utilisateur recherche [SOR 15]. Ce problème a été partiellement résolu par l'évolution des systèmes de recommandation [KAR 01] et l'existence de moteurs de recherche tels que Netflix, Amazon, YouTube et de nombreux autres sites sociaux ou de commerce électronique. Les systèmes de recommandation sont aujourd'hui devenus inévitables et leur tâche principale est de suggérer aux utilisateurs ce qu'ils devraient acheter, regarder, lire, etc., tout en tenant compte de leurs préférences et intérêts d'une manière pertinente. Ainsi, les éléments recommandés doivent en quelque sorte correspondre à ce que

l'utilisateur choisira réellement [SOR 15] en filtrant les informations non pertinentes que l'utilisateur ne sélectionnera pas [SIE 14]. Les systèmes de recommandation sont très importants pour améliorer et augmenter les revenus des entreprises d'une manière très subtile ; par conséquent, c'est une méthode intelligente pour rivaliser avec d'autres industries. En tant qu'exemple d'importance des systèmes de recommandation, il y a quelques années, Netflix a organisé un concours pour offrir 1 million de dollars à celui qui pourrait créer un algorithme de recommandation qui surpasse l'algorithme de Netflix [NET 09]. Du point de vue technique [SOR 15], l'objectif principal d'un système de recommandation est de suggérer des éléments pertinents aux utilisateurs en trouvant la corrélation entre les utilisateurs et certains éléments [SIE 14]. De nombreux algorithmes sont développés et utilisés pour créer des systèmes de recommandation, dont deux catégories sont régulièrement utilisées. Celles-ci correspondent au filtrage collaboratif et au filtrage basé sur le contenu. Cependant, la faible précision des suggestions faites par les systèmes de recommandation de filtrage collaboratif est devenue une grande préoccupation [VAR 18]. Il existe de nombreuses applications ou sites Web où les utilisateurs peuvent interagir de manière anonyme [KAR 01]. Par exemple, un utilisateur non enregistré peut toujours naviguer sur un site Web de boutique en ligne en ajoutant des articles à son panier pour les consulter plus tard. Autre exemple, un utilisateur non connecté peut tout de même regarder des vidéos sur YouTube. Nous devrions donc toujours être en mesure de recommander des articles à un utilisateur, que nous disposions ou non de données explicites à son sujet. Par conséquent, afin de savoir ce que l'utilisateur veut vraiment, nous devons savoir « pourquoi » l'utilisateur est en train de chercher quelque chose, nous devons connaître l'intention de l'utilisateur. La fouille de processus s'appuie sur les logs des activités pour découvrir les modèles de processus, vérifier leur conformité, améliorer les processus et recommander l'activité suivante. La fouille de processus peut donc être utilisée comme technique de recommandation pour suggérer la prochaine activité à l'utilisateur tout en s'appuyant sur le modèle de processus. Cependant, c'est une technique qui ne se concentre que sur les activités générées à partir d'un processus métier, elle est principalement orientée activité et ne traite pas la partie intentionnelle d'un processus. Les modèles intentionnels sont plus précis dans la représentation de l'intention de l'utilisateur et la détection du raisonnement humain derrière les activités produites. La fouille d'intentions devient un domaine de recherche important, dérivé de la fouille des processus et avec les mêmes objectifs, mais qui s'adresse aux modèles de processus intentionnels. Ainsi, nous voyons que la fouille d'intentions peut nous fournir une meilleure qualité de recommandation car elle se concentre sur la partie intentionnelle des processus qui est plus proche de la pensée et des objectifs de l'utilisateur.

L'objectif principal de la fouille d'intentions [KHO 13] est de connaître les objectifs et le but des utilisateurs lors de l'utilisation du système en découvrant les modèles de processus intentionnels qui reflètent le comportement et la stratégie de l'utilisateur. De plus, les processus réels mis en œuvre par les utilisateurs sont différents des processus métier établis. La principale raison de cette différence est que l'utilisateur ne s'appuie pas uniquement sur les processus prescrits pour accomplir ses activités quotidiennes, mais qu'il utilise également ses propres intentions et stratégies pour accomplir une certaine tâche. Pour aider l'utilisateur à atteindre son objectif et sa tâche de manière efficace, la fouille d'intentions doit être utilisée pour fournir une recommandation à l'utilisateur en lui donnant les connaissances nécessaires et en déterminant la séquence optimale d'actions afin d'atteindre son intention. Par conséquent, les logs du logiciel et les intentions déjà découvertes doivent être utilisés pour fournir une bonne recommandation à l'utilisateur. De l'autre côté, de nombreux facteurs environnementaux tels que le temps, l'emplacement, la météo et le profil sont obtenus à partir de nombreuses sources, telles que les capteurs, les systèmes externes, les acteurs externes ou des bases de connaissances du domaine, et pourraient également améliorer les recommandations. Les recherches existantes se concentrent principalement sur des ensembles de données de logs des activités uniquement ; seuls quelques-uns envisagent de combiner diverses sources. Notre travail explore comment combiner plusieurs types de sources avec la fouille de processus intentionnelle riche en contexte en utilisant des ontologies pour construire un système de recommandation basé sur la sémantique qui fournira une meilleure qualité de recommandations.

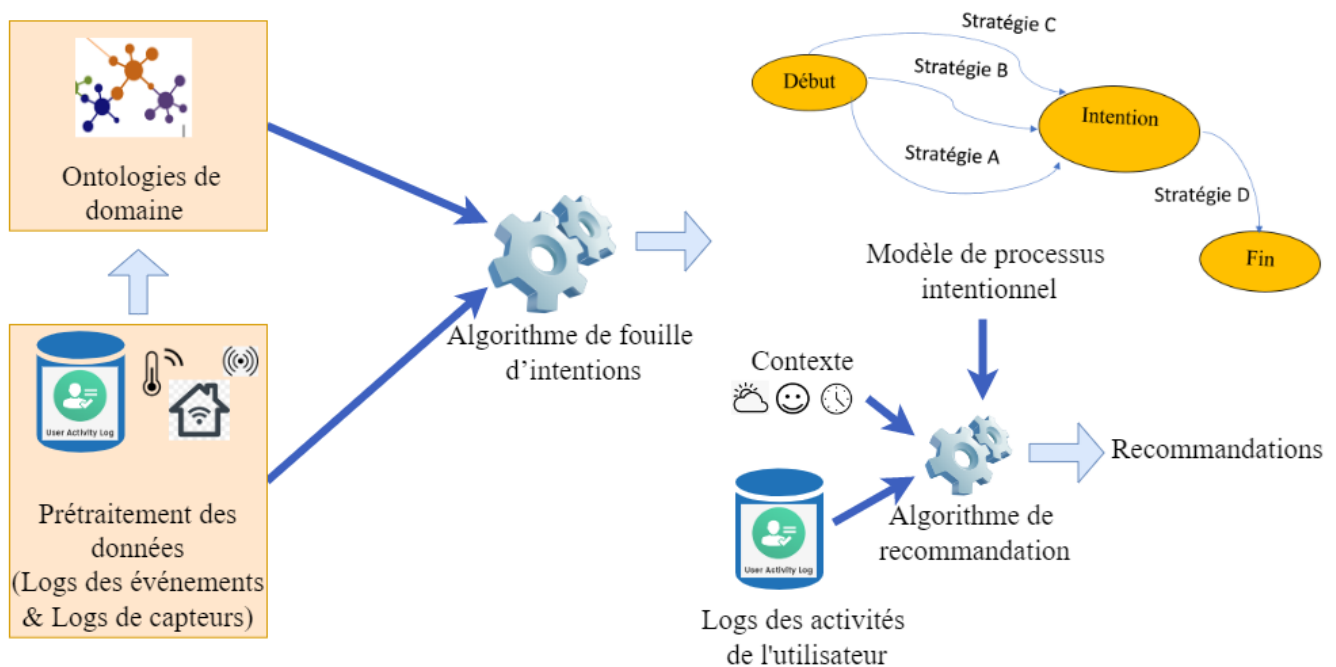


Figure 1. Approche proposée.

Comme les modèles intentionnels sont proches du raisonnement humain, nous pensons qu'en utilisant l'environnement contextuel pour construire ces modèles, il est possible de fournir des recommandations plus efficaces aux utilisateurs. En effet, le même utilisateur, dans des contextes différents, choisira des stratégies ou des intentions différentes pour aboutir à son objectif. Notre approche (cf. Figure 1) est la suivante : Au début, les différentes sources - logs des événements, logs de capteurs, etc. avec leurs informations contextuelles - sont collectées. Puis, on doit soit créer l'ontologie de domaine à partir des données existantes (dans le cas où celle-ci n'existe pas déjà) soit utiliser l'ontologie de domaine existante. Ensuite, un algorithme de fouille d'intentions construit un modèle intentionnel. Cet algorithme découvrira les intentions et les stratégies du modèle intentionnel à partir des logs d'activités, avec l'aide de l'ontologie de domaine. Le modèle créé utilisera le formalisme MAP [ROL 99]. Le modèle de processus intentionnel construit et les logs de l'utilisateur avec les informations contextuelles sont fournis comme paramètres d'entrée à l'algorithme de recommandation. Après avoir traité les paramètres d'entrée, l'algorithme de recommandation fournira des recommandations appropriées pour l'utilisateur en fonction de son activité actuelle.

2.2. Modèle de processus contextuel

Afin d'atteindre notre objectif, nous travaillons sur les données relatives à un cas de Smart Home puisque ce domaine est riche en informations contextuelles. Nous devons connaître les liens existants entre les différentes activités et les différents types de capteurs afin de pouvoir construire un modèle de processus contextuel qui pourra nous diriger dans la construction de notre modèle de processus intentionnel (cf. Figure 2).

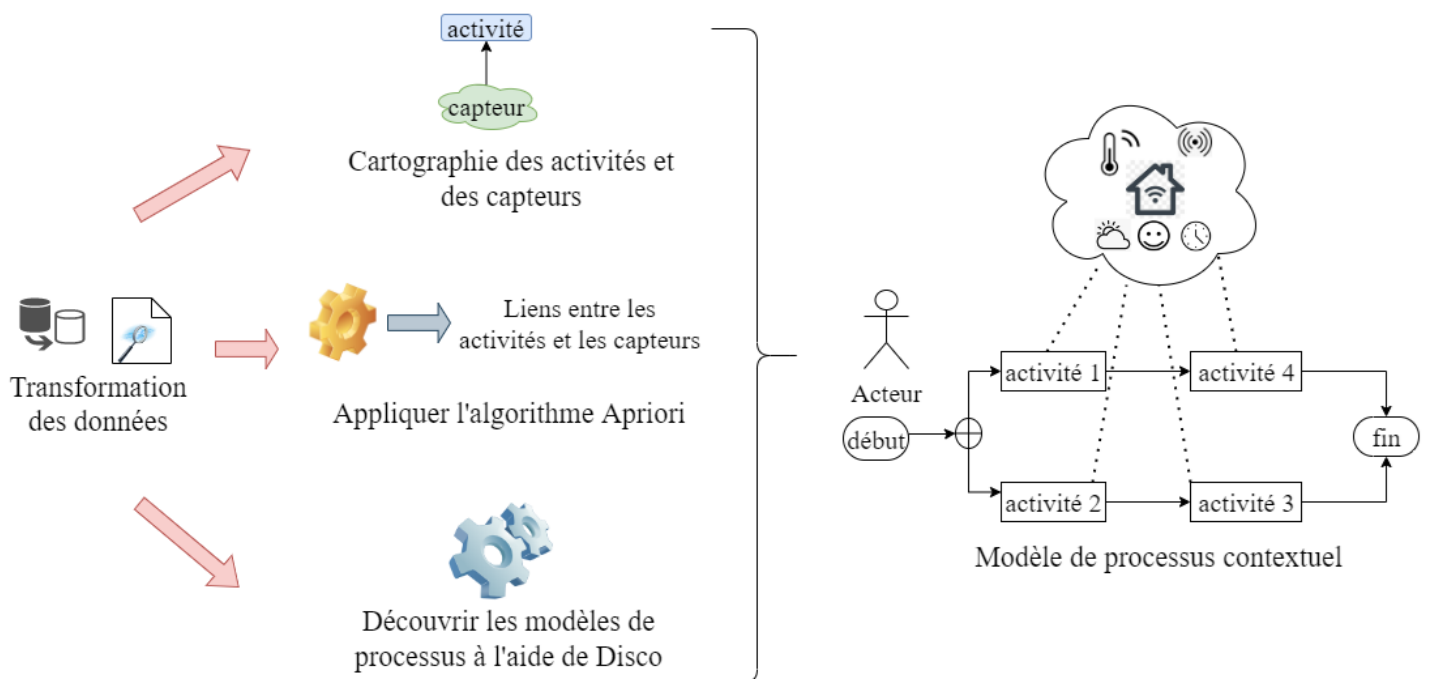


Figure 2. Approche pour construire le modèle de processus contextuel.

Afin de construire le modèle de processus contextuel, nous avons fait le lien entre les activités et les capteurs en utilisant l’algorithme Apriori [ZAK 14] et en générant le modèle de processus (en utilisant le logiciel Disco). Nous combinons les résultats des liens trouvés avec le modèle de processus obtenu afin de pouvoir annoter le modèle de processus par les informations contextuelles.

La Figure 3 présente un exemple du modèle de processus contextuel dans le cas d'une maison intelligente (Smart Home), où les activités du processus sont annotées par les informations contextuelles obtenues par les capteurs. L’indicateur « Confiance » indique le degré de confiance correspondant au fait que l’activité se déroule ou pas lorsque le capteur se déclenche (par exemple, il y a 81 % de chance que l’activité « Faire la vaisselle » se déroule lors du déclenchement du capteur « Capteur de vaisselle non lavée »). Le profil de l'acteur affecte également le choix des activités.

2.3. Perspectives

Après avoir trouvé le lien entre les différents types de capteurs et les activités, le but suivant est de construire l’algorithme de fouille d’intentions afin de pouvoir construire le modèle de processus intentionnel prenant en considération les informations contextuelles. Cela nous permettra de développer l’algorithme de recommandation utilisant les informations contextuelles, les activités en cours et le modèle de processus intentionnel pour présenter à la volée des recommandations à l’utilisateur.

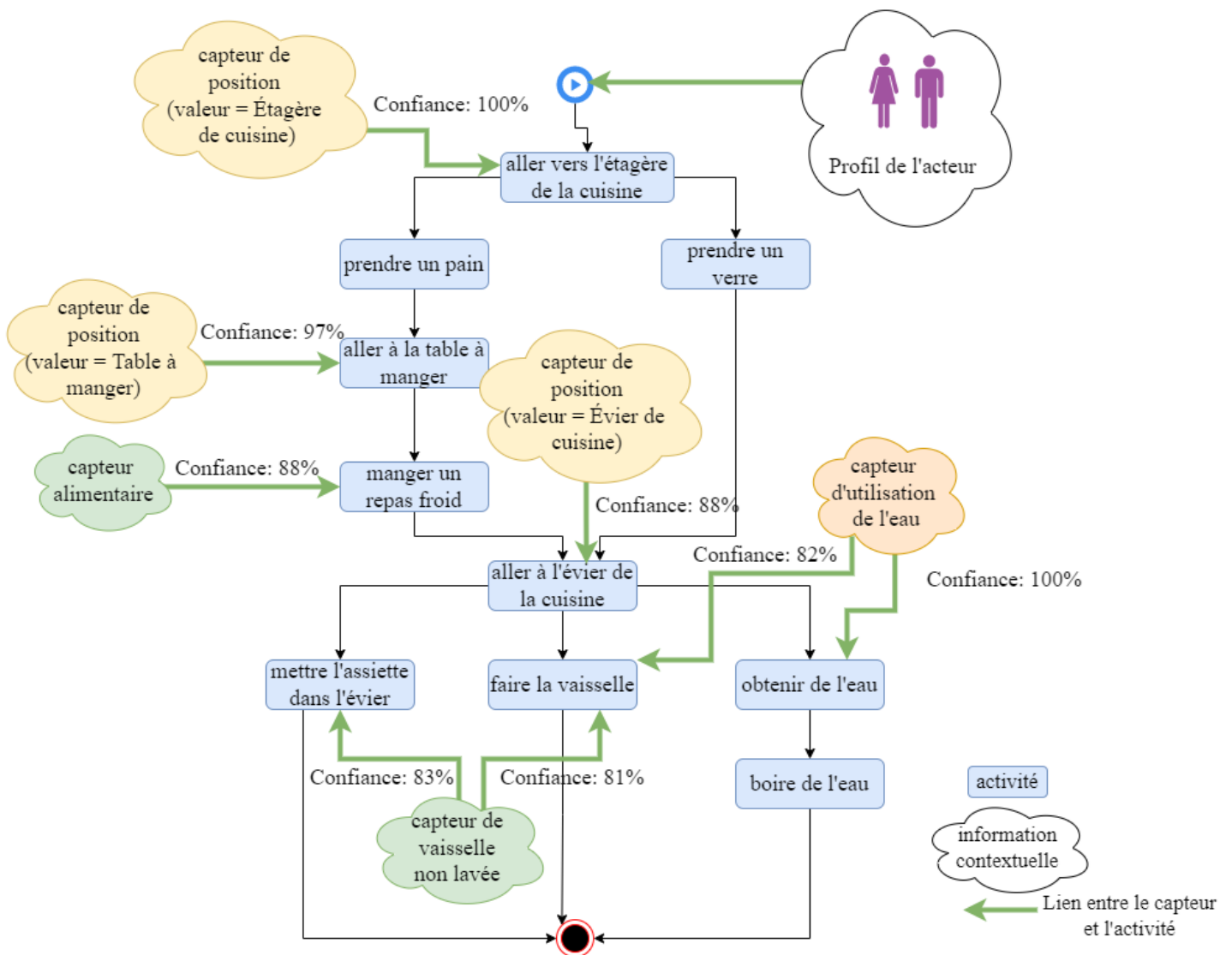


Figure 3. Exemple d'un modèle de processus contextuel.

3. Modélisation de l'origine biologique au service de l'étude de la chimio-diversité du vivant

La chimio-diversité du vivant est encore peu connue. Elle représente l'ensemble des composés d'origine naturelle, une bonne connaissance de cet ensemble permettrait pourtant une meilleure appréhension dans les domaines pharmacologiques et agroalimentaires répondant à des problématiques majeures. Plusieurs facteurs peuvent expliquer le manque de connaissances sur la diversité chimique, biologiques et/ou techniques, il faut en plus y ajouter l'érosion de la diversité biologique. Ce problème bien connu représente un défi temporel supplémentaire risquant de nous faire passer à côté de molécules, d'organismes d'intérêts. Cet ensemble de composés naturels est étudié grâce à la métabolomique qui permet d'avoir l'information la plus fidèle par rapport à ce qui se passe dans un système biologique.

L'un des freins techniques limitant cette connaissance a été mis en évidence par l'équipe de biologistes à l'origine de ce projet. En effet, les sources de données chimiques référencant les molécules connues sont spécifiques, elles répondent à des problématiques différentes les rendant incomplètes pour un traitement plus global. L'objectif ici est donc de réunir toutes les informations connues sur les composés d'origines biologiques, ainsi que leurs métadonnées dans une seule et même base de données exhaustive. Les données manquantes seront complétées grâce à des modèles prédictifs s'appuyant sur les données intégrées avant d'être confirmées par validations expérimentales [HEN 22].

3.1. Contexte

La métabolomique est l'étude des petits produits naturels, appelés métabolites, elle est le dernier maillon de la chaîne des « omiques » initiée par la génomique. Si le génome correspond à ce que l'organisme à la capacité de faire, l'analyse du métabolome permet d'appréhender un système biologique dans son environnement en captant ce qui se produit réellement. La métabolomique est donc un outil majeur de la compréhension du vivant. C'est aussi sur elle que repose les connaissances sur la chimio-diversité s'organisant autour des taxons, des milieux trophiques et des écosystèmes en une dynamique et une cartographie encore peu documentée et difficile à estimer [KIN 09]. Une meilleure connaissance de la chimio-diversité est un enjeu majeur dans plusieurs domaines tels que l'agroalimentaire ou encore en pharmacologie. Il est donc essentiel de trouver des solutions, et ce rapidement dans un contexte d'érosion de la biodiversité, afin de mettre en évidence des molécules, précurseurs ou encore organismes ou groupes d'organismes à fort potentiel. De plus, l'étude de la chimio-diversité pourrait représenter un point de vue nouveau pour l'étude sur l'évolution. Les perspectives sont encore larges, et son intérêt est confirmé dans la communauté scientifique, il faut donc en identifier les limites afin de réfléchir à des solutions.

À l'heure actuelle, pour identifier les molécules présentes dans un spectre à la sortie du spectromètre de masse (outil d'analyse en métabolomique) deux possibilités s'offrent aux biologistes. Comparer les fragments moléculaires obtenus ici avec des fragmentations déjà connues grâce à des bases de données expérimentales issues d'analyses sur des échantillons purs (que l'on appelle standard et qui sont peu nombreux au regard de toute la chimiodiversité car il s'agit d'un travail long, qui concerne donc une petite portion de composés modèles). Soit comparer les fragments à des fragmentations prédites à partir des formules moléculaires telles que *SMILES* (Simplified Molecular Input Line Entry Specification – langage symbolique de description de la structure des molécules chimiques) et *InChI* (International Chemical Identifier – Identifiant chimique international) de composés connus répertoriés dans des bases de données chimiques. En effet, certains logiciels spécialisés comme *MS-Dial* (Mass Spectrometry - Data Independent Acquisition – spectrométrie de masse - acquisition indépendante des données) [TSU 15] vont pouvoir prédire les cassures des molécules à partir des SMILES ou des InChI fournis par les bases en respectant les règles de chimie structurale.

Il n'existe pas de base de données chimique exhaustive, elles sont très spécialisées soit par classe chimique (*lipid Maps* pour les lipides) soit par branche taxonomique (*HMDB - Human Metabolome Database* - pour les métabolites humains). Il va donc falloir renouveler le traitement pour chaque base de données qui intéresse le biologiste chargé de l'expérimentation. C'est en travaillant sur un outil de traitement de données métabolomiques permettant l'amélioration de l'annotation des composés [FRA 20] que l'équipe du LRSV (Laboratoire de Recherche en Sciences Végétales - <https://www.lrsv.ups-tlse.fr/home/>) a mis en évidence une limite technique freinant l'analyse. En effet, en pré-sélectionnant les composés que l'on s'attend à trouver dans un échantillon, on réduit le temps de traitement, mais sans perdre en efficacité, au contraire même parfois. C'est à l'aide d'une première version de base de données unifiée que les biologistes ont pu faire ce constat, ils ont ensuite extrait de cette base de données les composés déjà référencés dans les sources de données chez l'espèce et la famille taxonomique de l'échantillon ainsi que les composés proches structurellement. Sur les 600 000 composés uniques de la base de données unifiée seulement un tiers est bio sourcé. On ne connaît l'origine biologique que d'une partie des composés. Certaines sources de données dont sont extraites les molécules ne contiennent même pas de champs dédiés à ce type de métadonnées dont nous avons pourtant confirmé l'intérêt pour l'amélioration des analyses métabolomiques, mais aussi permettant de cartographier la chimio-diversité.

Ce projet à donc pour but de construire un référentiel de données exhaustif de toutes les informations connues sur les produits naturels issues de plusieurs sources de données hétérogènes afin de développer des modèles prédictifs notamment pour prédire les origines biologiques des composés

de biosource inconnue. La validation sera faite expérimentalement à partir de divers échantillons biologiques afin de corriger et confirmer les modèles.

3.2. Vers la modélisation de l'origine biologique

Il n'existe pas encore à notre connaissance de travail réalisé sur la prédiction de l'origine taxonomique d'un composé naturel, néanmoins nous nous appuyés sur une étude mettant en évidence des représentations significatives de différentes classes chimiques dans les divers règnes du vivant [CHA 19].

Nous avons initié un premier travail sur une précédente version de la base de données unifiée afin de confirmer qu'il est effectivement possible de prédire l'origine biologique d'un composé. C'est engageant. Nos avancées sur la modélisation de l'origine biologique se poursuivent mais ne sont pas encore suffisamment significatives. Il reste un grand travail à produire sur l'homogénéisation des données.

Pour l'heure, un jeu de données réduit et équilibré a été constitué en respectant les contraintes suivantes ; chaque composé ne doit avoir qu'une origine biologique connue, et une structure chimique standardisée. Des données numériques caractérisant la structure chimique des composés ont été calculées. Les composés ayant un lien significatif avec l'origine biologique ont été ensuite utilisés dans la construction de modèles de clusterisation. L'objectif de ces modèles est d'identifier le règne taxonomique d'un composé à partir de données chimiques et structurales. Certains modèles ont donné des résultats satisfaisants tels qu'un modèle KNN – K-Nearest Neighbors (K plus proches voisins) ou encore un modèle Random Forest (forêt d'arbre décisionnel) ayant permis d'identifier correctement le règne pour 80% des composés testés (cf. Figure 4).

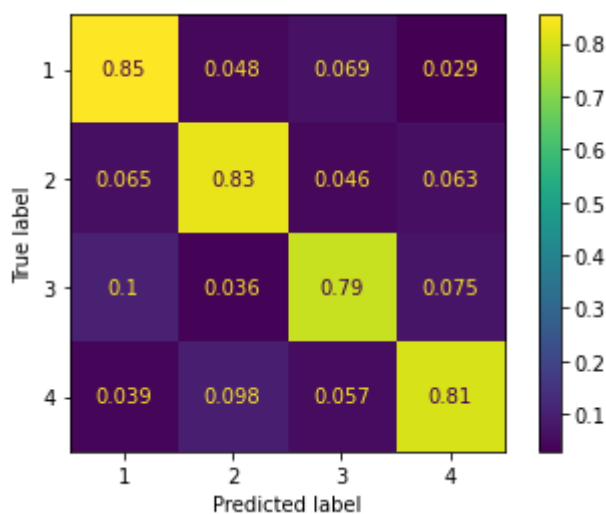


Figure 4. Résultat de prédiction du règne biologique avec un modèle random forest
1=Bactérie, 2=Champignon, 3=Animal, 4=Végétal.

Dans la réalité, des composés peuvent être issus de plusieurs sources biologiques, et certaines structures chimiques sont caractéristiques d'une branche taxonomique. C'est la cartographie de la chimio-diversité connue qui va nous permettre une meilleure appréhension de ces marqueurs. Nous n'avons pas à l'heure actuelle suffisamment de recul sur les données pour identifier clairement quel modèle pourra être mis en place pour modéliser de façon optimale la biosource des composés d'origine naturelle, ni même si un seul modèle permettra de tenir compte de la diversité du vivant. Nous devons tout d'abord finir l'état des lieux des informations dont nous pourrons disposer.

3.3. Perspectives

Une fois qu'une modélisation pertinente de la biosource d'un produit naturel aura été proposée, nous pourrions compter sur une collection d'organismes végétaux et macromycètes pour la validation. Avec les métadonnées sur les composés ainsi complétés, nous voudrions confirmer la plus-value d'un traitement métabolomique avec présélection d'un échantillon de composés d'intérêts, mais également de classement des priorités de l'annotation. Nous aurons à notre disposition différentes études en cours au sein de la plateforme de métabolomique hébergeant ce projet pour comparer cela avec des traitements actuellement utilisés de façon quotidienne en métabolomique. Nous voudrions également savoir s'il est possible d'aller plus loin en affinant la hiérarchisation de la classification chimique existante en biochimie et ainsi développer un modèle de classification chimique des composés en s'appuyant sur des travaux portant sur la même question, mais à l'aide des métadonnées de différentes classifications répertoriées dans des sources de données distinctes dont sont issus les composés que nous regroupons.

4. Modélisation d'un système de suivi et d'évaluation d'une activité de création de jeu vidéo en classe

Notre sujet porte sur la modélisation d'un système de suivi et d'évaluation d'une activité de création de jeu vidéo par des élèves de seconde en classe [VIG 22]. Créer un jeu apparaît comme un levier de motivation pour les élèves [REY 11]. Mais il apparaît très difficile d'évaluer l'impact de l'atelier sur les apprentissages scolaires et sur l'acquisition des compétences du 21^{ème} siècle telles que la collaboration ou la créativité [KAF 15]. L'évaluation de l'apprentissage dans un cadre scolaire suppose « mesurer le degré d'acquisition des connaissances et des compétences ainsi que la progression de l'élève » (Loi n°2013-595 du 8 juillet 2013 d'orientation et de programmation pour la refondation de l'école de la République). La compétence est définie comme la « capacité d'action efficace face à une famille de situations, qu'on arrive à maîtriser parce qu'on dispose à la fois des connaissances nécessaires et de la capacité à les mobiliser à bon escient » [PER 11].

4.1. Contexte

L'atelier de création de jeu vidéo est une « Situation d'Apprentissage Informatisée (SAI) », c'est-à-dire « une situation d'apprentissage intégrant un ou plusieurs logiciels qui y ont un rôle particulier du point de vue de l'apprentissage » [TCH 09]. Le logiciel central de l'activité est le moteur de création de jeu [WU 12] qui est la source d'information principale sur le travail effectué par les élèves à condition que l'enseignant puisse analyser des « traces d'apprentissage » [BOY 19]. Il s'agit donc de convertir des « traces numériques », « enregistrement automatique d'éléments d'interaction entre un utilisateur et son environnement, dans le cadre d'une activité spécifiée » [LAF 07] en « indicateurs », « variables au sens mathématique auxquelles sont attribuées une série de caractéristiques » [DIM 06]. A travers notre observation de l'activité dans quatre classes de seconde en France, nous avons pu identifier plusieurs freins. Tout d'abord, les moteurs de jeu ne proposent pas systématiquement de traces numériques liées à l'activité du créateur. Ensuite, les enseignants méconnaissent les processus de production d'un jeu et ne parviennent donc pas à effectuer le suivi à travers des évaluations formatives qui aident à identifier les forces et faiblesses de l'élève pour l'aider à progresser [AND 09]. Il est également difficile pour eux de faire le lien entre l'activité et les compétences visées, et donc de réaliser l'évaluation sommative qui reconnaît les objectifs d'apprentissage à atteindre par leurs étudiants [LEG 06]. Nous souhaitons proposer une solution pour résoudre ces difficultés en répondant aux deux questions suivantes : (1) comment identifier des données pertinentes à collecter puis les mettre en place ? et (2) comment traiter ces données pour en faire des informations significatives afin de suivre et organiser la progression des élèves ainsi que pour réaliser le suivi et l'évaluation de compétences (évaluations formatives et sommatives) ?

4.2. Modèle du suivi et d'évaluation en classe

Nous avons réalisé un modèle de données théorique permettant de répondre aux deux questions identifiées (cf. Figure 5). Pour commencer, nous nous appuyons sur des ouvrages de référence en conception des jeux (game design) et en production de jeu vidéo pour décrire le processus de réalisation d'un jeu vidéo [BJO 04] [SCH 14]. Un jeu se compose d'unités de conception ou concepts que les élèves souhaitent intégrer dans leur jeu (par ex : les personnages). Chaque unité de conception est décomposée en unité de production ou briques fonctionnelles à produire pour créer le jeu (ex : le personnage-joueur, les personnages non joueurs). Les élèves choisissent les fonctionnalités qu'ils souhaitent implémenter dans leurs jeux et réalisent les tâches de productions associées. Au fur et à mesure de l'avancement du projet, les unités de conception et de production passent donc par différents états, du prototype à l'état final. Ces changements d'état peuvent être tracés dans le moteur par des logs. Mais pour pouvoir identifier les informations significatives à traiter pour suivre et évaluer la progression des élèves, il faut pouvoir relier ces tâches de productions aux référentiels de compétences existants. Nous avons étudié le socle commun de compétences et les principes directeurs pour l'apprentissage du 21^{ème} siècle pour identifier les compétences transverses et les compétences scolaires, donc reliées à des disciplines. Cette étude a été complétée par une analyse des programmes scolaires de la classe de seconde qui identifie les connaissances à maîtriser pour un élève et par une lecture des manuels scolaires qui mettent en avant des exercices pour chaque connaissance.

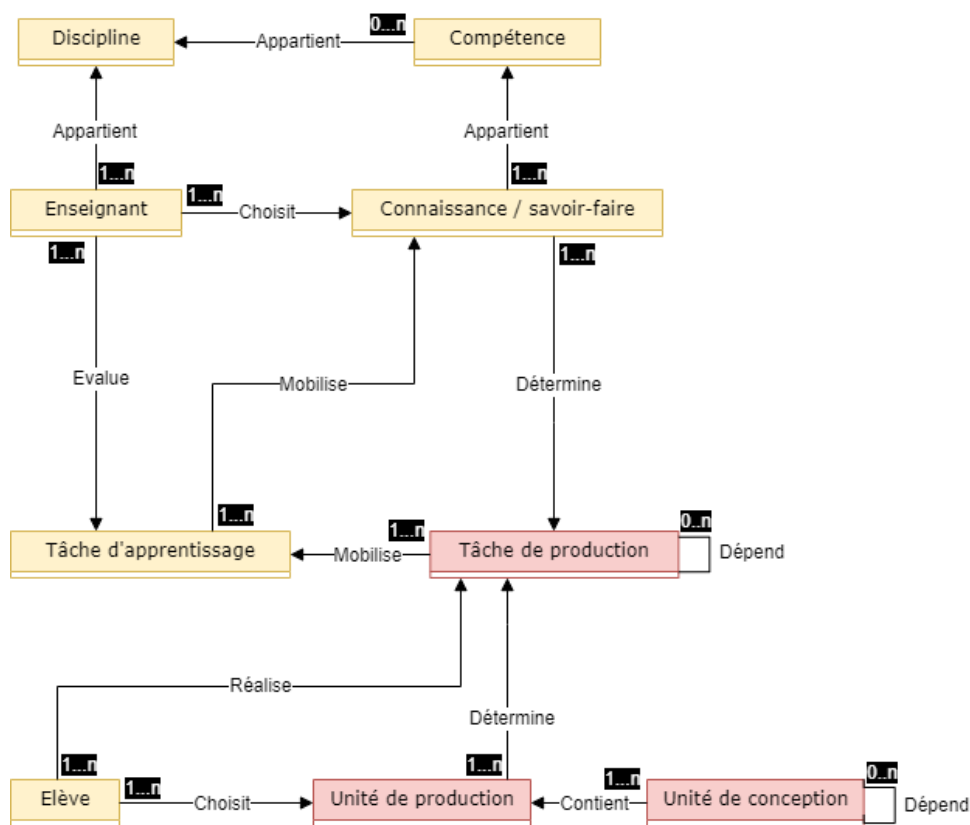


Figure 5. Modélisation du système de suivi et d'évaluation en classe.

Le concept commun entre la production d'un jeu vidéo et l'apprentissage scolaire est la notion de tâches. « La notion de tâche permet de lier l'activité de l'élève et les contenus tels qu'ils sont envisagés dans la programmation de l'enseignant. Elle permet d'envisager les apprentissages plus globalement que celle d'exercice » [REU 13]. Nous avons donc mis en lien les tâches de production d'un jeu vidéo et les tâches d'apprentissage pour concevoir le modèle ci-dessus.

4.3. Utilisation du modèle pour la création des jeux vidéo en classe

Le modèle de notre système de suivi et d'évaluation de l'activité de création de jeu vidéo est fondé sur le contrôle des changements d'état de chacune des unités de conception, donc sur la réalisation des

tâches de production correspondantes, elles-mêmes reliées à des tâches d'apprentissages qui mobilisent des connaissances. Pour contrôler ces changements d'états nous devons implémenter des données associées à chaque tâche qui fonctionneront comme des traces d'apprentissage. Ces données ne peuvent être vérifiées que dans un environnement maîtrisé. Nous proposons donc de développer des patrons de jeux composés d'un certain nombre d'unités de conception dans un état donné. Le passage d'un état à l'autre, donc la bonne réalisation de la tâche, peut être validé en s'appuyant sur les données implémentées dans ces patrons. Nous nous appuyons sur deux hypothèses de travail : (1) la vérification des tâches de production d'une phase à l'autre du projet va permettre de suivre la progression des élèves donc de réaliser des évaluations formatives et (2) la complétion de toutes les tâches d'apprentissage rattachées à une connaissance valide la compétence et sert donc de base à l'évaluation sommative.

Nous avons développé un premier patron de jeu narratif/énigmes sur le moteur *Construct3* dans lesquelles toutes les unités de production sont dans un état « terminé » sauf les unités « Personnages », « Histoire » et « Son » qui sont dans un état « brouillon » (cf. Figure 6). Ce patron de jeu a pour objectif de faire travailler les compétences d'expression orale et écrite, de développement de sa créativité, de résolution de problèmes, d'utilisation des outils numériques, d'honnêteté intellectuelle (à travers les notions de droits d'auteur sur les images et sons) et d'organisation de son travail. Il est actuellement utilisé en classe par des enseignants de Français, d'Histoire-Géographie, de Science du Numérique, de Physique ou encore d'Arts plastiques. Les élèves peuvent réécrire tous les textes, changer les images des personnages, et modifier les sons (les énigmes proposées sont basées sur l'écoute des fréquences radio et d'appels). La modification de texte s'effectue simplement en supprimant le texte existant et le remplaçant par le nouveau texte rédigé en classe. Pour changer les images des personnages, les élèves doivent remplacer dans le moteur l'url de l'image intégrée dans le patron par l'url d'une image qu'ils ont choisie. Cela implique donc de chercher des images libres de droit. Un emplacement est prévu pour intégrer les crédits d'images et de son qui sont également exportés dans un fichier. Toute modification de son ou de texte peut entraîner un changement des bonnes réponses à trouver pour résoudre les énigmes. Les variables des énigmes sont donc modifiables également par les élèves.



Figure 6. Patron de jeu narratif / énigmes.

Nous avons implémenté un système de logs pour suivre la réalisation des tâches de production. Un bouton permet d'exporter tous les textes rédigés dans un fichier csv, ce qui facilite la relecture pour l'enseignant de Français par exemple. Le même fonctionnement est valable pour les crédits d'image et de son ou encore les variables d'indice qui indiquent si l'élève a modifié les énigmes. Pendant et après la séance de cours, il est donc possible de suivre les tâches de production réalisées ou à faire. Ces logs servent aussi de base pour évaluer individuellement les tâches d'apprentissage.

Une grille d'évaluation globale des jeux est également fournie. Elle sert à effectuer l'évaluation sommative de l'activité. Elle aide l'enseignant à mesurer le degré de personnalisation du jeu et à visualiser l'ensemble des tâches de production et leur lien avec les tâches d'apprentissage et les connaissances / compétences associées. Cette grille est destinée à aider l'enseignant à apprécier le travail de manière générale et à appréhender la création de jeu vidéo comme une activité interdisciplinaire.

Dans les deux modalités, les enseignants restent donc maîtres de leur évaluation finale mais ils peuvent s'appuyer sur les données issues des réalisations des élèves, soit par tâches, soit dans la vue globale de la grille d'évaluation.

4.4. Perspectives

Ce premier patron de jeu propose un nombre limité de compétences / connaissances à mobiliser et donc de tâches d'apprentissage à travailler. Les expérimentations menées actuellement sont destinées à vérifier deux éléments : (1) la pertinence du modèle conçu qui fait le lien entre compétences travaillées et activité de création de jeu vidéo et (2) le potentiel du patron de jeu pour suivre et évaluer l'activité. L'évaluation de ces expérimentations s'appuie sur des questionnaires élèves et enseignants, des observations en classe et une analyse des données recueillies en jeu.

Les premiers retours des enseignants sont favorables mais montrent des limites au niveau des élèves qui réalisent les tâches de production sans pour autant influencer la structure du jeu. Pour permettre de travailler des compétences transverses telles que la créativité ou le travail collaboratif, le patron proposé doit donc proposer un plus grand nombre de tâches et laisser aux élèves le choix entre des tâches plus diversifiées dans la limite des compétences que les enseignants souhaitent travailler. Il serait donc nécessaire de pouvoir produire un patron plus modulaire.

De plus, lors du développement de ce premier patron, nous avons été confrontés aux limites techniques du moteur de jeu Construct3 pour générer des logs. Nous avons donc réalisé un comparatif des autres moteurs de jeu qui pourraient être utilisés dans un cadre pédagogique tout en permettant la mise en place de ces « traces d'apprentissage ». Nous réalisons actuellement des tests sur le moteur open source Godot.

A la suite de la première expérimentation et en fonction du résultat, nous souhaitons donc essayer de produire un second patron de jeu sur un nouveau moteur et réaliser une nouvelle série d'expérimentations en classe.

5. Conclusion

Cet article résume trois travaux de recherche des doctorantes en première ou en deuxième année de thèse pour l'année universitaire 2021-2022. Les auteures ont présenté leurs thèses lors du Forum JCJC de la conférence INFOSRID en juin 2022 et ont été sélectionnées en tant que meilleures contributions parmi les douze travaux des JCJC de cette année. Les doctorantes ont retravaillé leurs articles suivant la perspective modélisation.

Les trois travaux de thèse très différents démontrent bien la nécessité de construire les modèles conceptuels. Le développement de nouvelles technologies basées sur les données massives et

l'apprentissage automatique ne remettent pas en cause les pratiques de modélisation, bien au contraire, justifient ce travail d'abstraction. Par conséquent, les pratiques de modélisation évoluent pour prendre en compte la nature plus dynamique des données et des modèles, la multiplicité des sources d'information, l'aide de l'apprentissage automatique et de l'intelligence artificielle. Les opportunités offertes par les technologies émergentes sont nombreuses et vont permettre de s'adapter à de nouveaux usages et de nouvelles expériences utilisateurs.

Bibliographie

- [AND 09] ANDRADE H.L., CIZEK, G.J., *Handbook of formative assessment*, Routledge, 2009.
- [BJO 04] BJORK S., HOLOPAINEN J., *Patterns in game design*, Charles River Media, 2004.
- [BOY 19] BOYER A., « Quelques réflexions sur l'exploration des traces d'apprentissage », *Distances et médiations des savoirs*, n°27, 2019.
- [CHA 19] CHASSAGNE, F., CABANAC, G., HUBERT, G., DAVID, B., & MARTI, G., « The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products® », *Phytochemistry Reviews*, volume 18, p.601-622, 2019.
- [DIM 06] DIMITRACOPOULOU A., BRUILLARD E., « Enrichir les interfaces de forums par la visualisation d'analyses automatiques des interactions et du contenu », *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, n°13, p. 354-397, 2006.
- [ELA 22] ELALI, R., « Recommandations contextuelles fondées sur la fouille d'intentions et les ontologies », *Actes du 11^{ème} Forum des Jeunes Chercheuses Jeunes Chercheurs de la conférence INFORSID*, 2022.
- [FET 09] FETTKE P., « How Conceptual Modeling Is Used », *Communications of the Association for Information Systems*: Vol. 25, Article 43, 2009.
- [FRA 20] FRAISIER-VANNIER, O., CHERVIN, J., CABANAC, G., PUECH, V., FOURNIER, S., DURAND, V., AMIEL, A., ANDRÉ, O., BENAMAR, O. A., DUMAS, B., TSUGAWA, H., & MARTI, G., « MS-CleanR : A Feature-Filtering Workflow for Untargeted LC-MS Based Metabolomics », *Analytical Chemistry*, volume 92, p.9971-9981, 2020.
- [HAR 20] HÄRER F., FILL H.-G., « Past Trends and Future Prospects in Conceptual Modeling - A Bibliometric Analysis », *International Conference on Conceptual Modeling*, Springer, 2020.
- [HEN 22] HENNECHART, S., « Intégration de données et inférence au service de l'étude de la chimiodiversité du vivant », *Actes du 11^{ème} Forum des Jeunes Chercheuses Jeunes Chercheurs de la conférence INFORSID*, 2022.
- [KAF 15] KAFAI Y., QUINN B., « Constructionist Gaming Understanding the Benefits of Making Games for Learning », *Educational Psychologist*, n°50, vol 4, p.313-334, 2015.
- [KAR 01] KARYPIS, G., « Evaluation of item-based top-n recommendation algorithms », *International Conference on Information and Knowledge Management*, ACM, 2001.
- [KHO 13] KHODABANDELOU, G., HUG, C., DENECKERE, R., SALINESI, C., « Process Mining Versus Intention Mining », *Enterprise, Business-Process and Information Systems Modeling*, pp. 466-480, Spain, Springer, 2013.
- [KIN 09] KIND, T., SCHOLZ, M., & FIEHN, O., « How Large Is the Metabolome? A Critical Analysis of Data Exchange Practices in Chemistry ». *PLoS ONE*, volume 4, p. e5440, 2009.
- [LAF 07] LAFLAQUIERE J., PRIE Y., « Des traces modélisées, un nouveau support pédagogique ? », *Actes de la 4e conférence scientifique de Lornet*, 2007.
- [LEG 06] LEGENDRE R, *Dictionnaire actuel de l'éducation*, Éditions Guérin, 2006.
- [NET 09] "NETFLIX PRIZE" Netflixprize.Com, www.netflixprize.com. 2009.
- [PER 11] PERRENOUD P., *Construire des compétences dès l'école. Pratiques et enjeux pédagogiques* ESF Sciences humaines, 2011.
- [REC 21] RECKER J.C., LUKYANENKO R., JABBARI M., SAMUEL B.M., Castellanos A., « From representation to mediation: a new agenda for conceptual modeling research in a digital world », *MIS Quarterly: Management Information Systems*, 45(1), pp. 269-300, 2021.

- [REU 13] REUTER Y., HAREL COHEN-AZRIA C., DAUNAY B., DELCAMBRE I., LAHANIER-REUTER D., *Dictionnaire des concepts fondamentaux de didactique*, Editions De Boeck Supérieur, 2013.
- [REY 11] REYNOLDS R., HAREL CAPERTON I., « Contrasts in student engagement, meaning-making, dislikes, and challenges in a discovery-based program of game design learning », *Education Tech Research Dev*, n°59, p.267-289, 2011.
- [ROL 99] ROLLAND C., PRAKASH N., BENJAMEN A., « A multi-model view of process modelling », *Requirements Engineering Journal*, 4(4), pp. 169-187, 1999.
- [SCH 14] SCHELL J., *The Art of Game Design: A Book of Lenses*, A K Peters/CRC Press, 2014.
- [SIE 14] SIELIS, G. A., TZANAVARI, A., PAPADOPOULOS, G. A., « Recommender Systems Review of Types, Techniques, and Applications », *Encyclopedia of Information Science and Technology*, Third Edition, 2014.
- [SOR 15] SORDE, R. K., DESHMUKH, S. N., « Comparative Study on Approaches of Recommendation System », *International Journal of Computer Applications*, Volume 118–No. 2, 2015.
- [TCH 09] TCHOUNIKINE P., *Précis de recherche en ingénierie des EIAH*, <https://lig-membres.imag.fr/tchounikine/Articles/PrecisV1.pdf>, 2009.
- [TSU 15] TSUGAWA, H., CAJKA, T., KIND, T., MA, Y., HIGGINS, B., IKEDA, K., KANAZAWA, M., VANDERGHEYNST, J., FIEHN, O., ARITA, M., « MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis », *Nature Methods*, volume 12, p.523-526, 2015.
- [VAR 18] VARZANEH, H. SOLEIMANI NEYSIANI, B., ZIAFAT, H., SOLTANI, N., « Recommendation Systems Based on Association Rule Mining », *Emerging Science Journal*, 2(2), 2018.
- [VIG 22] VIGNEAU C., « Conception d'un système d'évaluation d'une activité de création de vidéo pour les enseignants de seconde », *Actes du 11^{ème} Forum des Jeunes Chercheuses Jeunes Chercheurs de la conférence INFORSID*, 2022.
- [WAN 02] WAND Y., WEBER R., « Research Commentary: Information Systems and Conceptual Modeling—A Research Agenda », *Information Systems Research* (13)4, pp. 363–377, 2002.
- [WU 12] WU B., WANG A.I., « A guideline for Game Development Based Learning », *International Journal of Computer Games Technology*, vol 12, 20p., 2012.
- [ZAK 14] ZAKI, J., M., MEIRA J., W., *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press, 2014.