

Création Semi-Automatisée d'Ontologies chez Lettria

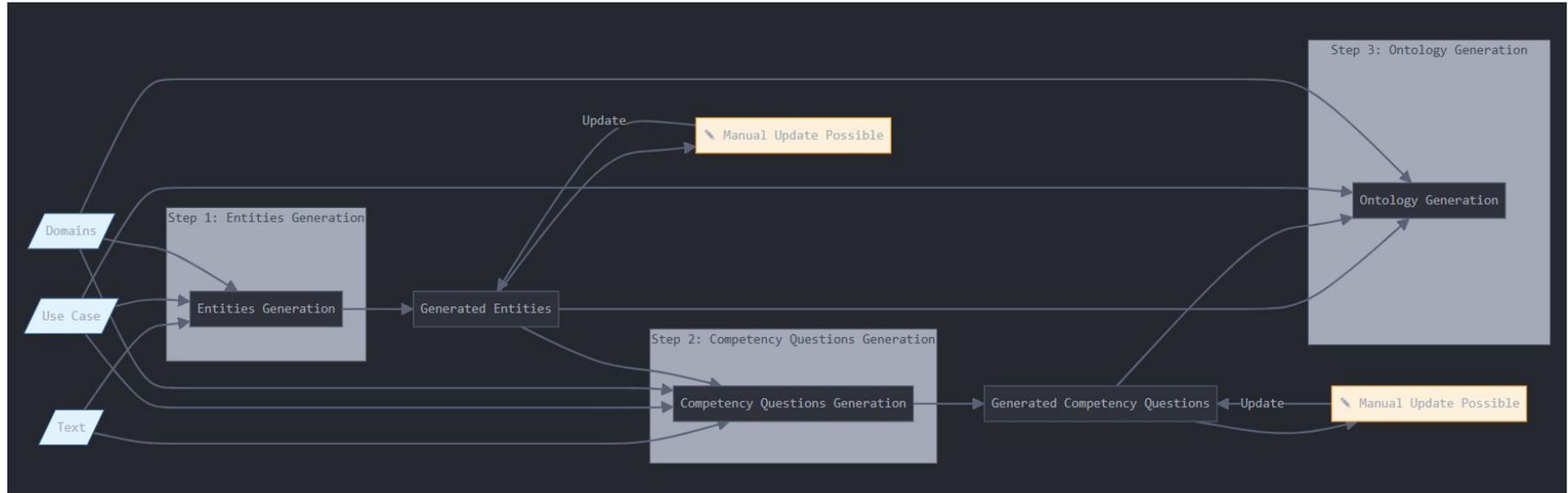
Julien Plu, Oscar Moreno Escobar, Edouard Trouilleux,
Axelle Gapin and Raphaël Troncy



julien@lettria.com

@julienplu

Ontology-Toolkit : Pipeline



- Ce pipeline représente notre façon de générer une ontologie en indiquant les parties modifiables
- 3 étapes :
 - Génération des type d'entités (classes)
 - Génération de questions de compétences
 - Génération de l'ontologie (classes, propriétés et axiomatisation)

Ontology-Toolkit : Génération des Types d'Entités

- Entrée :
 - Un texte
 - Un ou plusieurs domaines de compétences (ex: littérature, cinéma, médical, etc...)
 - Un cas d'utilisation : utiliser pour diriger le modèle afin que la génération soit la plus adaptée possible à comment on souhaite l'utiliser et pourquoi
- Sortie :
 - Une liste de type d'entités identifié dans le texte et correspondant au domaine et au cas d'utilisation
- Intervention utilisateur :
 - Possibilité d'ajouter / modifier / supprimer les types générés afin de corriger ou compléter la sortie du modèle

Ontology-Toolkit: Génération Des Questions de Compétences

- **Entrée :**
 - Un texte
 - Un ou plusieurs domaines de compétences (ex: littérature, cinéma, médical, etc...)
 - Un cas d'utilisation : utiliser pour diriger le modèle afin que la génération soit la plus adaptée possible à comment on souhaite l'utiliser et pourquoi
 - Les types d'entités générés
- **Sortie :**
 - Une liste de questions de compétences utilisant les types comme "placeholder" donnant une direction des propriétés à prendre en compte lors de la génération de l'ontologie
- **Intervention utilisateur :**
 - Possibilité d'ajouter / modifier / supprimer les questions générées afin de corriger ou compléter la sortie du modèle

Ontology-Toolkit: Génération de l'Ontologie

- Entrée :
 - Un ou plusieurs domaines de compétences (ex: littérature, cinéma, médical, etc...)
 - Un cas d'utilisation : utiliser pour diriger le modèle afin que la génération soit la plus adaptée possible à comment on souhaite l'utiliser et pourquoi
 - Les types d'entités générés
 - Les questions de compétences générées
- Sortie :
 - Une ontologie respectant les types et les questions précédemment générés

Evaluation Quantitative et Qualitative

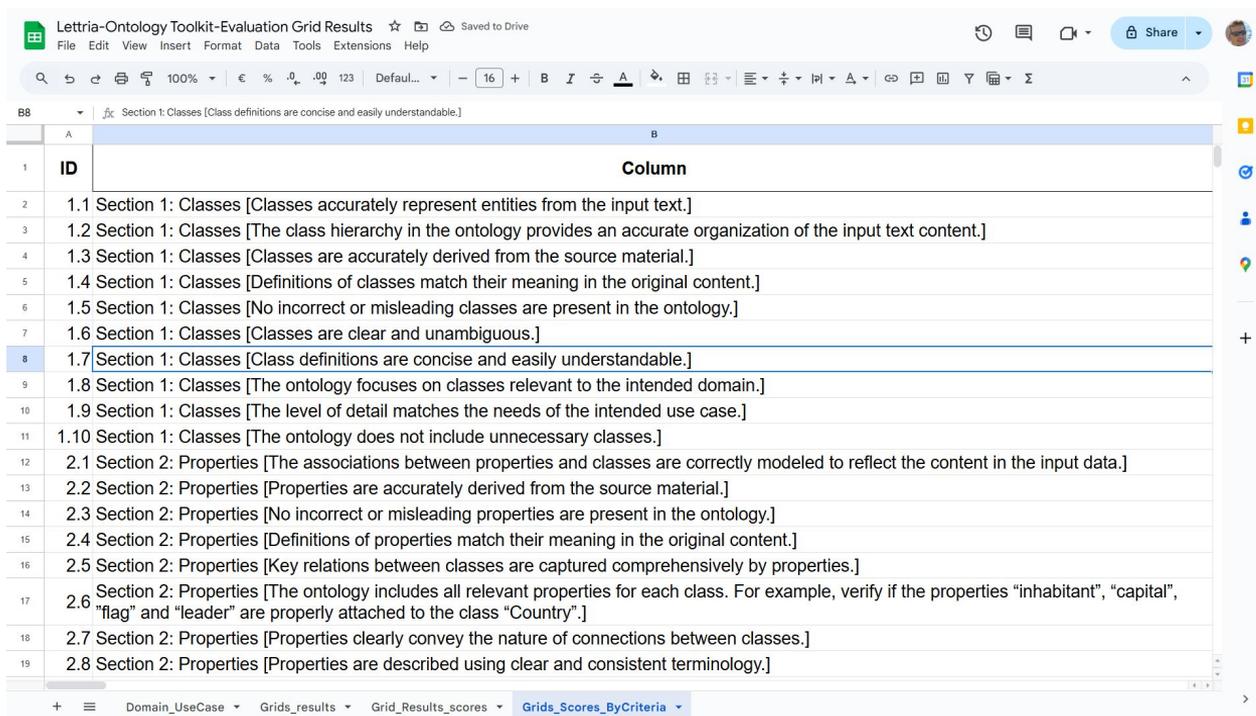
- Evaluation quantitative :
 - sur une ontologie de référence dans le domaine de la finance :
 - nombre de classes : 37
 - nombre de propriétés : 102
 - génération de 12 ontologies ayant chacun leur configuration :
 - nombre de questions : 10, 30, 50 choisi de manière empirique
 - avec / sans cas d'utilisation
 - avec / sans auto-correction
 - 3 LLMs : Claude 3.5 Sonnet v1, GPT4o 05-2024, GPT4o-mini 05-2024
- Evaluation qualitative :
 - utilisation d'une grille d'évaluation
 - 8 participants, chacun sur un domaine différents
 - utilisation de la meilleure config

Evaluation Quantitative : Résultats

Model	Ontology	Type	Yes	Narrower	Broader	No	Total Yes, Narrower, Broader
Claude 3.5 Sonnet	With use case 50	classes	13 (35.14%)	2 (5.41%)	11 (29.73%)	11 (29.73%)	70.27%
		properties	25 (24.51%)	0 (0.00%)	43 (42.16%)	34 (33.33%)	66.67%
	Without use case 50	classes	13 (35.14%)	5 (13.51%)	10 (27.03%)	9 (24.32%)	75.68%
		properties	15 (14.71%)	0 (0.00%)	44 (43.14%)	43 (42.16%)	57.84%
GPT-4o	With use case 50	classes	13 (35.14%)	1 (2.70%)	11 (29.73%)	12 (32.43%)	67.57%
		properties	5 (4.90%)	7 (6.86%)	0 (0.00%)	90 (88.24%)	11.76%
	Without use case 50	classes	5 (13.51%)	1 (2.70%)	16 (43.24%)	15 (40.54%)	59.46%
		properties	2 (1.96%)	7 (6.86%)	38 (37.25%)	55 (53.92%)	46.08%
GPT-4o-mini	With use case 50	classes	12 (32.43%)	1 (2.70%)	13 (35.14%)	11 (29.73%)	70.27%
		properties	1 (0.98%)	0 (0.00%)	37 (36.27%)	64 (62.75%)	37.25%
	Without use case 50	classes	15 (40.54%)	0 (0.00%)	12 (32.43%)	10 (27.03%)	72.97%
		properties	5 (4.95%)	0 (0.00%)	17 (16.83%)	79 (78.22%)	21.78%

- Meilleures configs:
 - 50 questions
 - sans auto-correction
 - avec cas d'utilisation : meilleures précision sur les propriétés
 - sans cas d'utilisation : meilleures précision sur les classes
 - Claude Sonnet 3.5 v1

Evaluation Qualitative : Grille d'évaluation



The screenshot shows a Google Sheet with the following table content:

	A	B
1	ID	Column
2	1.1	Section 1: Classes [Classes accurately represent entities from the input text.]
3	1.2	Section 1: Classes [The class hierarchy in the ontology provides an accurate organization of the input text content.]
4	1.3	Section 1: Classes [Classes are accurately derived from the source material.]
5	1.4	Section 1: Classes [Definitions of classes match their meaning in the original content.]
6	1.5	Section 1: Classes [No incorrect or misleading classes are present in the ontology.]
7	1.6	Section 1: Classes [Classes are clear and unambiguous.]
8	1.7	Section 1: Classes [Class definitions are concise and easily understandable.]
9	1.8	Section 1: Classes [The ontology focuses on classes relevant to the intended domain.]
10	1.9	Section 1: Classes [The level of detail matches the needs of the intended use case.]
11	1.10	Section 1: Classes [The ontology does not include unnecessary classes.]
12	2.1	Section 2: Properties [The associations between properties and classes are correctly modeled to reflect the content in the input data.]
13	2.2	Section 2: Properties [Properties are accurately derived from the source material.]
14	2.3	Section 2: Properties [No incorrect or misleading properties are present in the ontology.]
15	2.4	Section 2: Properties [Definitions of properties match their meaning in the original content.]
16	2.5	Section 2: Properties [Key relations between classes are captured comprehensively by properties.]
17	2.6	Section 2: Properties [The ontology includes all relevant properties for each class. For example, verify if the properties "inhabitant", "capital", "flag" and "leader" are properly attached to the class "Country".]
18	2.7	Section 2: Properties [Properties clearly convey the nature of connections between classes.]
19	2.8	Section 2: Properties [Properties are described using clear and consistent terminology.]

- 30 critères :
 - 10 sur les classes
 - 11 sur les propriétés
 - 9 sur l'aspect globale
- Note entre 1 et 5 pour chaque critère

Evaluation Qualitative : Résultats

Users	S1: Classes	S2: Properties	S3: Overall	Average
1	4	4	3.7	3.9
2	3.1	3.7	2.1	3
3	3.8	4.4	4.1	4.1
4	4.7	4.6	3.7	4.3
5	4.1	3.9	3.8	3.9
6	5	5	5	5
7	3.8	3.7	4	3.8
8	3.4	4.2	3.6	3.7
Average All Users	3.99	4.19	3.74	3.97

- Les retours indiquent que :
 - L'ontologie générée offre une conception forte et bien structurée
 - Sa conception est souvent peu profonde et plate, et montre des limites
 - Trop de classes de haut niveau et une profondeur insuffisante dans la hiérarchie

Demo

Ontology-Toolkit Demo

Select Model

Claude 3.5 Sonnet

Generate Classes

Fill out the form below. Once you submit, we will generate classes based on the document, use case and domain you provide.

Document

The Sega Genesis, known as the Mega Drive outside North America, is a 16-bit fourth generation home video game console developed and sold by Sega. It was Sega's third console and the successor to the Master System. Sega released it in 1988 in Japan as the Mega Drive, and in 1989 in North America as the Genesis. In 1990, it was distributed as the Mega Drive by Virgin Mastertronic in Europe.

Domain

Video Game Console

Use Case

Modeling Video Game Consoles characteristics and history in order to better create a complete graph ab

Generate

Conclusion et Travail à Venir

- Résultats encourageants
- Déjà utilisé :
 - pour nos clients
 - en interne pour le démarrage d'un projet autour de la génération automatique de graphes de connaissances
- Des pistes d'améliorations :
 - perfectionner les prompts
 - gérer la hiérarchisation des classes et des propriétés
 - gérer la profondeur des classes et des propriétés (abstrait vs concret)
 - gérer l'incrémentation d'une ontologie
 - gérer l'alignement avec d'autres ontologies
- Envie d'en savoir plus ?
 - Notre site : <https://www.lettria.com/>
 - Contact : charles@lettria.com