

# M1 Informatique – UE Projet

## Carnet de bord v2 : les coulisses de la recherche documentaire

*Les éléments que vous indiquez dans ce carnet donneront lieu à une notation*

Noms, prénoms et spécialité :

Boumessaoud Abdelkader STL
Abdellaoui Zaky STL

Sujet :

Benchmarking de solutions optimistes pour génération de données test à partir de JSON-Schema

Consigne :

1. **Introduction (5-10 lignes max)** : Décrivez rapidement votre sujet de recherche, ses différents aspects et enjeux, ainsi que l'angle sous lequel vous avez décidé de le traiter.
2. **Les mots clés retenus** : Listez les mots-clés que vous avez utilisés pour votre recherche bibliographique. Organisez-les sous forme de carte heuristique.
3. **Descriptif de la recherche documentaire (10-15 lignes)** : Décrivez votre utilisation des différents outils de recherche (moteurs de recherche, base de donnée, catalogues, recherche par rebond etc.). Comparez ensuite les outils entre eux. A quelles sources vous ont-ils permis d'accéder ? Quelles sont leurs spécificités ? Quel est leur niveau de spécialisation ?
4. **Bibliographie produite dans le cadre du projet** : Utilisez la norme ACM.
5. **Evaluation des sources (5 lignes minimum par source)** : Choisissez 3 sources parmi votre bibliographie, décrivez la manière dont vous les avez trouvées et faites-en une évaluation critique en utilisant les critères vus sur les supports de TDs.

## Introduction :

Le projet de recherche dont il est question consiste à étudier différentes approches pour générer automatiquement des exemples de données à partir de la structure des données JSON. Pour ce faire, les chercheurs utilisent un langage d'assertions appelé JSON-Schema qui permet de décrire la structure des données JSON.

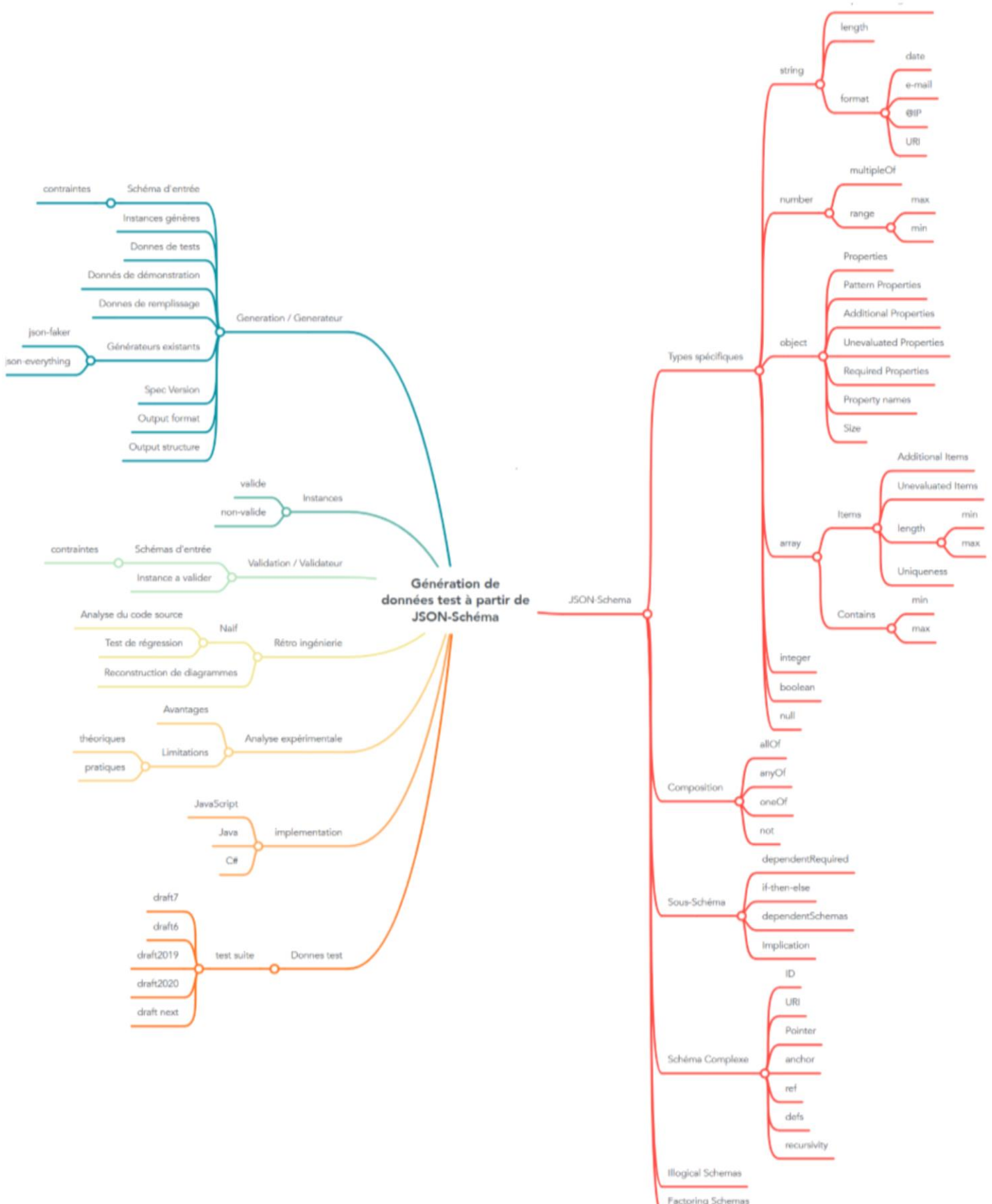
Le but de cette étude est de trouver des méthodes efficaces pour générer des exemples de données de test qui pourront être utilisés pour vérifier si les données JSON que l'on reçoit sont conformes à la structure définie dans le JSON-Schema.

Il existe plusieurs méthodes pour générer ces exemples de données de test, mais certaines peuvent produire des données non conformes, ce qui peut poser des problèmes dans certains contextes.

Le projet se concentre sur l'étude de trois bibliothèques open-source qui prennent en compte la quasi-totalité des opérateurs du langage de schéma. L'objectif est de comprendre les limites de ces approches, en combinant une analyse rétro-ingénierie et une analyse expérimentale utilisant des exemples de données JSON réels et synthétiques.

Pour mener à bien ce projet, il est nécessaire d'avoir des connaissances en langages informatiques tels que Java, JavaScript et C#, ainsi qu'un sens analytique et une aptitude à formaliser des problèmes ayant une finalité pratique.

## Les mots clés retenus :



## Descriptif de la recherche documentaire :

Lors de notre recherche documentaire, ont utilisé plusieurs outils de recherche tels que Google Search, Google Scholar, Wikipédia, GitHub, Stack Overflow, ACM Digital Library, IEEE Xplore, DBLP, arXiv et Semantic Scholar.

**Google Search** : est un moteur de recherche généraliste qui fournit des résultats sur une grande variété de sujets, mais les résultats peuvent être trop nombreux et ne pas être spécifiques à des besoins particuliers.

**Google Scholar** : est plus spécialisé et se concentre sur les publications scientifiques, y compris les articles de revues, les thèses et les livres. Cet outil est utile pour trouver des résultats précis et pertinents dans le domaine scientifique.

**Wikipédia** : est une encyclopédie en ligne qui fournit des informations générales sur divers sujets. Les articles sont écrits par des contributeurs bénévoles et peuvent varier en termes de qualité et de fiabilité.

**GitHub** : est une plateforme de développement de logiciels qui permet d'accéder à des projets open source liés à un sujet de recherche, y compris du code, des problèmes et des discussions sur des projets particuliers.

**Stack Overflow** : est une plateforme de questions-réponses dédiée à la programmation informatique et utile pour trouver des solutions à des problèmes spécifiques liés à la programmation.

**ACM Digital Library, IEEE Xplore et DBLP** : sont des bibliothèques numériques qui donnent accès à des articles de revues, des conférences et des livres dans le domaine de l'informatique. ACM Digital Library couvre un large éventail de domaines liés à l'informatique, tandis qu'IEEE Xplore fournit également des normes. DBLP répertorie des articles de revues, des actes de conférences et des chapitres de livres spécifiquement dans le domaine de l'informatique.

**arXiv** : est une archive en ligne de prépublications scientifiques dans les domaines des mathématiques, de la physique, de l'informatique et de l'astrophysique.

**Semantic Scholar** : est un moteur de recherche académique qui utilise l'intelligence artificielle pour extraire des informations clés des articles scientifiques et les rendre facilement accessibles et interprétables.

En somme, chaque outil a ses avantages et inconvénients et peut être utilisé de manière complémentaire pour obtenir une recherche complète et fiable dans le domaine de l'informatique et des sciences.

## Bibliographie produite dans le cadre du projet :

- [1] A. Antonio, « Exploitez des données au format JSON », *OpenClassrooms*.  
<https://openclassrooms.com/fr/courses/7697016-creez-des-pages-web-dynamiques-avec-javascript/7911021-exploitez-des-donnees-au-format-json>.
- [2] L. Attouche, M.-A. Baazizi, D. Colazzo, G. Ghelli, C. Sartiani, et S. Scherzinger, « Witness Generation for JSON Schema », *Proc. VLDB Endow.*, vol. 15, n° 13, p. 4002-4014, sept. 2022, doi: 10.14778/3565838.3565852.
- [3] M.-A. Baazizi, D. Colazzo, G. Ghelli, et C. Sartiani, « Counting types for massive JSON datasets », in *Proceedings of The 16th International Symposium on Database Programming Languages*, New York, NY, USA, sept. 2017, p. 1-12. doi: 10.1145/3122831.3122837.
- [4] M.-A. Baazizi, D. Colazzo, G. Ghelli, et C. Sartiani, « Schemas and Types for JSON Data: From Theory to Practice », in *Proceedings of the 2019 International Conference on Management of Data*, New York, NY, USA, juin 2019, p. 2060-2063. doi: 10.1145/3299869.3314032.
- [5] M.-A. Baazizi, D. Colazzo, G. Ghelli, C. Sartiani, et S. Scherzinger, « An Empirical Study on the “Usage of Not” in Real-World JSON Schema Documents », in *Conceptual Modeling: 40th International Conference, ER 2021, Virtual Event, October 18–21, 2021, Proceedings*, Berlin, Heidelberg, oct. 2021, p. 102-112. doi: 10.1007/978-3-030-89022-3\_9.
- [6] M.-A. Baazizi, D. Colazzo, G. Ghelli, C. Sartiani, et S. Scherzinger, « Negation-Closure for JSON Schema ». arXiv, 27 février 2022. doi: 10.48550/arXiv.2202.13434.
- [7] M.-A. Baazizi, D. Colazzo, G. Ghelli, et C. Sartiani, « Parametric schema inference for massive JSON datasets », *VLDB J. — Int. J. Very Large Data Bases*, vol. 28, n° 4, p. 497-521, mars 2022, doi: 10.1007/s00778-018-0532-7.
- [8] M. Droettboom, « Understanding JSON Schema », [En ligne]. Disponible sur: <https://json-schema.org/understanding-json-schema/UnderstandingJSONSchema.pdf>
- [9] A. A. Frozza, R. dos S. Mello, et F. de S. da Costa, « An Approach for Schema Extraction of JSON and Extended JSON Document Collections », in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, juill. 2018, p. 356-363. doi: 10.1109/IRI.2018.00060.
- [10] M. Fruth, M.-A. Baazizi, D. Colazzo, G. Ghelli, C. Sartiani, et S. Scherzinger, « Challenges in Checking JSON Schema Containment over Evolving Real-World Schemas », in *Advances in Conceptual Modeling: ER 2020 Workshops CMAI, CMLS, CMOMM4FAIR, CoMoNoS, EmpER, Vienna, Austria, November 3–6, 2020, Proceedings*, Berlin, Heidelberg, nov. 2020, p. 220-230. doi: 10.1007/978-3-030-65847-2\_20.
- [11] C. Köhnen, *Witness Generation for JSON Schema Patterns*. Gesellschaft für Informatik e.V., 2023. doi: 10.18420/BTW2023-79.
- [12] T. Lv, P. Yan, et W. He, « Survey on JSON Data Modelling », *J. Phys. Conf. Ser.*, vol. 1069, p. 012101, août 2018, doi: 10.1088/1742-6596/1069/1/012101.
- [13] F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, et D. Vrgoč, « Foundations of JSON Schema », in *Proceedings of the 25th International Conference on World Wide Web*, Montréal Québec Canada, avr. 2016, p. 263-273. doi: 10.1145/2872427.2883029.

## Evaluation des sources :

**Source [8]** : « Understanding JSON Schema » de M. Droettboom; Trouvable sur [Google Scholar](#) fournit un guide pratique pour apprendre à utiliser JSON Schema de manière efficace. Le livre couvre la spécification JSON Schema Draft 2020-12 et fournit des exemples pratiques ainsi que des notes spécifiques aux langages de programmation pour aider les lecteurs à comprendre les concepts plus facilement. Le ton amical et décontracté du livre en fait une lecture agréable pour les débutants et les programmeurs expérimentés. Les instructions sont claires et bien structurées, ce qui facilite l'apprentissage de la validation de la structure des données JSON. En somme, c'est un excellent document pour les personnes qui souhaitent apprendre et utiliser JSON Schema.

**Source [12]** : « Survey on JSON Data Modelling » de T. Lv, P. Yan et W. He. Trouvable sur [Semantic Scholar](#) fournit une vue d'ensemble des modèles de données JSON. Il examine les avantages et les inconvénients des différentes méthodes de modélisation des données JSON et les compare aux formats de données structurés traditionnels tels que XML. L'article présente également des exemples d'utilisation de différentes méthodes de modélisation des données JSON pour la gestion des données, la visualisation et l'analyse de données. Dans l'ensemble, l'article est bien structuré et fournit une introduction solide aux modèles de données JSON. Cependant, étant donné que l'article a été publié en 2018, il peut ne pas couvrir les dernières avancées en matière de modélisation de données JSON.

**Source [13]** : « Foundations of JSON Schema » de F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, et D. Vrgoč; Trouvable sur [ACM Digital Library](#) présente une analyse détaillée de la syntaxe et de la sémantique de JSON Schema, qui est un langage de validation pour les données JSON. L'article offre une perspective approfondie sur la façon dont JSON Schema peut être utilisé pour décrire et valider la structure de données JSON, ainsi que pour faciliter l'interopérabilité entre différents systèmes. En résumé, cet article est une lecture utile pour ceux qui s'intéressent à la modélisation de données JSON et à la validation des données.