



HAL
open science

Actes de la 3ème journée : Technologies du langage humain et accès interactif à l'information (avril 2022)

Benjamin Piwowarski, Florian Boudin, Gaël Dias, Jean-Pierre Chevallet, Jose G. Moreno, Mathieu Roche, Thierry Charnois

► To cite this version:

Benjamin Piwowarski, Florian Boudin, Gaël Dias, Jean-Pierre Chevallet, Jose G. Moreno, et al.. Actes de la 3ème journée : Technologies du langage humain et accès interactif à l'information (avril 2022). 3ème journée : Technologies du langage humain et accès interactif à l'information (2022), pp.1-36, 2022. hal-03946132

HAL Id: hal-03946132

<https://hal.science/hal-03946132>

Submitted on 19 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



AfIA

Association française
pour l'Intelligence Artificielle



GDR Groupement
de recherche
TAL Traitement automatique
des langues

Journée N° 3

Technologies du Langage Humain et Accès Interactif à l'Information

Collège TLH de l'AfIA, ARIA
et GdR TAL du CNRS

■ Préface

La journée «Accès interactif à l'information» organisée conjointement par le Collège Technologies du Langage Humain (TLH) de l'AFIA (AFIA), par l'Association Francophone de Recherche d'Information et Applications (ARIA) et le Groupement de recherche en Traitement Automatique de la Langue (GDR TAL), a pour objectif de proposer un panorama des recherches réalisées par les laboratoires francophones sur la thématique du traitement du langage humain et la recherche d'information, ainsi que ses applications liées à l'accès à l'information. Ainsi, nous avons souhaité réunir des collègues issus d'instituts de recherche menant des travaux à l'intersection du traitement automatique des langues, de la recherche d'information et de l'intelligence artificielle.

Suite à un appel à communications, nous avons reçu 13 contributions. La diversité des recherches présentées ainsi que la qualité et la quantité des contributions reçues démontrent une dynamique importante des TLH dans la communauté francophone. Ainsi, nous avons pu articuler l'organisation de la journée autour de cinq sessions thématiques : agents conversationnels, recherche interactive de documents, systèmes de questions-réponses, recommandation et simplification, et accès à l'information pour les non-voyants/malentendants. Ces actes rassemblent les résumés des communications issues de cette journée.¹

La journée est également l'occasion d'assister à deux présentations invitées :

- Sophie Rosset, Directrice de Recherche du CNRS, LISN : Méthodes d'évaluation des systèmes de dialogue : un tour d'horizon
- Lynda Tamine, Professeure Université Paul Sabatier, IRIT : Des tâches de recherche simples aux tâches de recherche complexes : Tirer parti de l'interaction en recherche d'information

Par ailleurs, un échange sur les enjeux et défis de l'accès interactif à l'information est également organisé en rassemblant les invités qui seront accompagnés de chercheurs industriels et académiques que nous remercions :

- Ludovic Denoyer - Facebook - FAIR / Sorbonne Universités
- Cédric Lopez - Emvista
- Catherine Pelachaud - CNRS / ISIR, UPMC
- Charles Teissèdre - Synapse Développement

Nous sommes ravis d'avoir pu réunir à cette occasion la communauté scientifique intéressée par l'accès à l'information. Nous remercions à ce titre le GDR-TAL, l'ARIA, l'AFIA, le comité scientifique, les intervenants et invités et le public.

Benjamin Piwowarski, Florian Boudin, Gaël Dias, Jean-Pierre Chevallet, José G. Moreno, Mathieu Roche, et Thierry Charnois.

1. Du fait de l'évolution de la situation sanitaire, la journée qui devait initialement accueillir 50 personnes a été remportée en 2022.

Table des matières

Adaptive conversational agent using reinforcement learning	1
<i>Lucie Galland, Florian Pecune, Catherine Pelachaud</i>	
CATI : système d'exploration et annotation assistée de grandes collections de documents	3
<i>Cédric Boscher, Elöd Egyed-Zsigmond</i>	
Développement de méthodes de question/réponse pour l'accès interactif à des collections de documents d'archives dans le cadre du projet ANR ARCHIVAL	5
<i>Frederic Bechet, Elie Antoine, Jeremy Auguste, Geraldine Damnati</i>	
Does Structure Matter? Leveraging Data-to-Text Generation for Answering Complex Information Needs	7
<i>Hanane Djeddal, Thomas Gerald, Laure Soulier, Karen Pinel-Sauvagnat, Lynda Tamine</i>	
Ero.Dot, un agent conversationnel pour le théâtre par combinaison de connaissances discrètes et continues	9
<i>Ilyes Bentebib, Jean-Pierre Chevallet, Didier Schwab</i>	
Evaluation des systèmes de recherche d'information interactifs : exemple du Bouclage de Pertinence implicite par suivi oculaire	14
<i>Philippe Mulhem, Francis Jambon, Lucas Albarede</i>	
Interactive Query Clarification and Refinement via User Simulation	17
<i>Pierre Erbacher, Ludovic Denoyer, Laure Soulier</i>	
Les projets TactiNET et TagThunder : quelle modalité pour un skimming et un scanning non visuels de l'information textuelle?	20
<i>Fabrice Maurel</i>	
Litte_Bot : le bot qui donne la réplique dans le style Molière	22
<i>Anna Pappa</i>	
Recherche du but de l'utilisateur par une session interactive	24
<i>Agnès Mustar, Sylvain Lamprier, Benjamin Piwowarski</i>	
Renewal - plateforme d'évaluation online pour la recommandation d'articles d'actualité	27
<i>Julien Hay, Bich-Liên Doan, Fabrice Popineau, Anne-Catherine Letournel</i>	
SigLa - Une plateforme de développement d'animations en langue des signes	29
<i>Bastien David, Jonathan Mutal, Irene Strasly, Johanna Gerlach, Pierrette Bouillon</i>	
Simplification automatique de textes scientifiques : SimpleText Lab à CLEF-2022	32
<i>Liana Ermakova, Sílvia Araújo, Patrice Bellot, Radia Hannachi, Stéphane Huet, Jaap Kamps, Elise Mathurin, Diana Nurbakova, Irina Ovchinnikova, Eric Sanjuan</i>	
Un agent conversationnel pour améliorer la découvrabilité des données ouvertes	34
<i>Anthony Gigerich, Mathieu Morey</i>	

Adaptive conversational agent using reinforcement learning

Lucie Galland

Florian Pécune

Catherine Pelachaud

In this work, we train our agent using deep RL to optimize its conversational strategies based on users' expected level of W and C of the agent. To maximize user's engagement, our agent adapts its conversational strategies at each of its speaking turn by selecting the appropriate dialog act (either linked to the task or to a social goal) to be perceived by the user with their preferred level of W and C . At each speaking turn, the simulated user module approximates the behavior of real users and computes which intentions to convey; for this it generates a dialog act (Stolcke *et al.*, 1998) and a sequence of non-verbal behaviors. The agent then relies on its engagement estimator to infer the simulated user's perceived level of engagement from the latter's non-verbal behavior and turn-taking information, as in (Sidner & Dzikovska, 2002). This estimation updates the dialog state. Depending on how the simulated user reacted to the agent's previous action, the agent updates the estimated user type using the type estimator module. Next, the agent's dialog manager produces a dialog act based on the current dialog state and the user's estimated type. Finally, the simulated user updates its own engagement according to the agent's performed dialog act, and produces another dialog act that updates the dialog state and gives back the turn to the agent.

For our model to adapt its strategies to the user, it needs to encounter different types of users. The user types are used to cluster all the possible preferences of the users. Since W and C account almost entirely for how people characterize others (Fiske *et al.*, 2007), we define our user types along these two dimensions and design 3 types of users : (WC) (respectively (W), (C)) user expects the agent to be W and C (respectively W , C).

The dialog manager is a Deep Q Neural network (DQN). A DQN is a neural network that takes as input the state of the agent and outputs Q values over the possible actions. The next action is chosen using a ϵ -greedy policy. The network is optimized to maximize the reward. The state of our dialog manager is :

$$S = \{ \text{user type, agent act, user act, mean of the last 3 engagement values, number of turns,} \\ \text{estimated topic engagement, historic of the number of time each strategy was used} \}$$

The action space is composed of the 7 dialog acts and the reward is crafted to balance task and social rewards. The task reward relates to the transmission of information while the social reward corresponds to the engagement maximization goal and coherent conversational strategies.

The dialog manager was trained in interaction with our simulated user. After training, the average engagement of the simulated user is higher when interacting with our adaptive model than when interacting with a non-adaptative baseline. The adaptive model is significantly improving the engagement of users expecting C . We also conducted a user study where participants were presented with videos of the interaction between our Greta ECA playing the role of a museum guide and a simulated user acting as a visitor. We observed that the adaptation is perceivable by the participants. The agent facing a user expecting W was judged significantly warmer than the baseline. The agent facing a user expecting only competence was found on average more competent than the baseline.

Références

FISKE S., CUDDY A. & GLICK P. (2007). Universal dimensions of social cognition : Warmth and competence. *Trends in cognitive sciences*, **11**, 77–83. DOI : [10.1016/j.tics.2006.11.005](https://doi.org/10.1016/j.tics.2006.11.005).

SIDNER C. & DZIKOVSKA M. (2002). Human-robot interaction : engagement between humans and robots for hosting activities. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, p. 123–128. DOI : [10.1109/ICMI.2002.1166980](https://doi.org/10.1109/ICMI.2002.1166980).

STOLCKE A., SHRIBERG E., BATES R., COCCARO N., JURAFSKY D., MARTIN R., METEER M., RIES K., TAYLOR P., VAN ESS-DYKEMA C. *et al.* (1998). Dialog act modeling for conversational speech. In *AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, p. 98–105.

CATI : système d'exploration et d'annotation assistée de grandes collections de documents

Cédric Boscher

Előd Egyed-Zsigmond

Université de Lyon, LIRIS UMR 5250 CNRS

cedric.boscher@insa-lyon.fr

elod.egyed-zsigmond@insa-lyon.fr

Mots clés : Recherche d'information, Interface Humain-Machine, Classification Interactive, Active Learning

CATI (Bosetti & Egyed-Zsigmond, 2020) est un outil d'exploration et d'annotation de grandes collections de documents (plusieurs centaines de milliers potentiellement). Il est à destination d'utilisateurs experts métier et/ou non-initiés à la Data Science et au Traitement Automatique du Langage Naturel. L'enjeu de nos travaux est d'offrir une solution simple de prise en main permettant d'extraire des informations à partir de grands volumes de documents, il est principalement motivée par les intérêts potentiels d'acteurs économiques et d'experts métiers dans le cadre de projets spécifiques, tels que le projets IDENUM¹ et LIVRONS² qui s'intéressent à une représentation visuelle et à grande échelle de la ville, à travers l'analyse de publication de textes et de photos sur les réseaux sociaux.

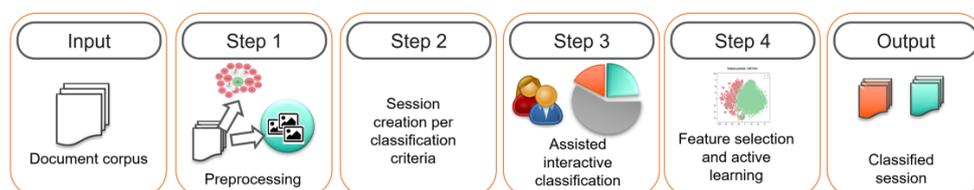


FIGURE 1 – Description de l'approche de classification implémentée dans CATI

Comme le montre la figure 1, CATI implémente une approche d'exploration et de classification de documents en plusieurs étapes.

La première étape consiste en l'import d'un jeu de données brut de documents texte pouvant intégrer des images (dans notre cas, des tweets ou des articles de journaux) dans CATI. À partir de ce jeu de données, nous créons dans un second temps des sessions, à savoir une instance du corpus importé, et voué à être exploré et annoté, en fonction de la pertinence de chaque document selon un critère de classification défini (par exemple, un laboratoire d'économie des transports cherchant à extraire les tweets ayant un rapport avec l'activité cyclo-logistique urbaine).

1. <https://imu.universite-lyon.fr/projet/idenum-identites-numeriques-urbaines/>

2. <https://imu.universite-lyon.fr/projet/livrons-livraison-a-velo-representations-sociales-et-donnees-des-reseaux-sociaux-2020/>

La troisième étape de l’approche consiste en l’exploration et l’annotation du jeu de données au moyen d’une analyse qualitative des données. Utilisant une base de données orientée documents (Elasticsearch) et des outils de traitement du langage naturel (SpaCy, NLTK) et des images, l’outil offre une grande flexibilité dans la recherche de documents, en s’appuyant à la fois sur les métadonnées des documents et sur des formes textuelles. Notre outil s’appuie sur des métaphores visuelles et interactives permettant d’explorer et d’annoter des collections de documents textes accompagnés d’images (dans notre cas, des tweets et des articles de presse) à grande échelle et en très peu de clics. Nous fournissons pour cela des fonctionnalités et métaphores visuelles telles que des clusters d’images similaires (Gaillard *et al.*, 2018), la reconnaissance d’objets dans des images (Redmon & Farhadi, 2018), la visualisation de N-Grams et de nuages de mots populaires, la visualisations de tendances hebdomadaires ainsi qu’un module de détection d’événements basée sur des anomalies de fréquence de termes (Guille & Favre, 2014).

Enfin, nous implémentons un module de classification automatique de corpus basé sur des méthodes d’Active Learning, afin de permettre à un utilisateur de classifier un grand corpus à partir de l’annotation d’un petit échantillon de documents. Nous offrons une interface visuelle permettant de guider l’utilisateur dans le choix des features à prendre en compte pour la classification.

Dans le cadre de notre présentation de CATI, nous proposerons une exploration sommaire des fonctionnalités de base de CATI ainsi que des métaphores visuelles faisant l’originalité et la simplicité de prise en main de notre approche, suivi d’une démonstration d’un cas d’usage réel s’appuyant sur un scénario de classification de jeu de données via l’interface de CATI.

Références

- BOSETTI G. & EGYED-ZSIGMOND E. (2020). Cati : An extensible platform supporting assisted classification of large datasets. In A. BOZZON, F. J. DOMÍNGUEZ MAYO & J. FILIPE, Éd.s., *Web Information Systems and Technologies*, p. 127–147, Cham : Springer International Publishing.
- BOSETTI G., EGYED-ZSIGMOND E. & ONO L. (2019). CATI : An Active Learning System for Event Detection on Mibroblogs’ Large Datasets :. In *Proceedings of the 15th International Conference on Web Information Systems and Technologies*, p. 151–160, Vienna, Austria : SCITEPRESS - Science and Technology Publications. DOI : [10.5220/0008355301510160](https://doi.org/10.5220/0008355301510160).
- GAILLARD M., EGYED-ZSIGMOND E. & GRANITZER M. (2018). CNN features for Reverse Image Search. *Document Numérique*, **21**(1-2), 63–90. DOI : [10.3166/RIA.21.1-31](https://doi.org/10.3166/RIA.21.1-31), HAL : [hal-02014019](https://hal.archives-ouvertes.fr/hal-02014019).
- GUILLE A. & FAVRE C. (2014). Mention-anomaly-based Event Detection and Tracking in Twitter. In *2014 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining*, Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining, p. 375–382, Beijing, China. DOI : [10.1109/ASONAM.2014.6921613](https://doi.org/10.1109/ASONAM.2014.6921613), HAL : [hal-01075294](https://hal.archives-ouvertes.fr/hal-01075294).
- REDMON J. & FARHADI A. (2018). Yolov3 : An incremental improvement, <https://pjreddie.com/darknet/yolo/>.

Développement de méthodes de question/réponse pour l'accès interactif à des collections de documents d'archives dans le cadre du projet ANR ARCHIVAL

Frederic Bechet¹, Elie Antoine¹, Jeremy Auguste¹, Géraldine Damnati²

(1) Aix Marseille Université, CNRS, LIS UMR 7020 prenom.nom@lis-lab.fr

(1) Orange Labs, Lannion prenom.nom@orange.com

Cette présentation se place dans le cadre du projet ANR ARCHIVAL¹ (FMSH – Université Paris 8 ; IRISA Rennes ; LIS – Marseille ; Orange Labs – Lannion). ARCHIVAL vise à faire collaborer des chercheurs issus des Sciences de l'Information et de la Communication et de l'Informatique autour de la valorisation des archives et du partage des savoirs pour les arts, la culture et le patrimoine. Il s'est structuré autour des questionnements suivants :

- Quel rôle peuvent jouer les méthodes de compréhension par les machines dans la réinterprétation de fonds d'archives thématiques ?
- Selon quelles modalités des interfaces de médiation des contenus peuvent-elles exploiter des résultats générés par les méthodes actuelles d'Intelligence Artificielle ?

Archival s'inscrit dans la lignée des projets exploitant les technologies du langage pour les Humanités Numériques et en particulier pour la curation de documents comme par exemple (Rehm *et al.*, 2019), en proposant de surcroît de nouveaux modes d'interaction pour l'accès à des collections archivistiques :

- l'exploration progressive par le biais de liens calculés automatiquement, soit en amont, soit dynamiquement en fonction d'une zone d'intérêt sélectionnée par l'utilisateur
- l'interrogation explicite de la collection par le biais d'une question

Parmi les différentes méthodes envisagées dans le projet pour générer des liens entre documents et interagir avec ceux-ci, nous nous intéressons spécifiquement dans cette communication aux méthodes de compréhension du langage de type *Machine Reading Comprehension* que nous exploitons pour les deux modes d'interaction ci-dessus.

Paradigme question/réponse pour l'accès interactif à l'information - Le but de cette étude est d'évaluer si les avancées récentes en compréhension de lecture (ou MRQA pour *Machine Reading Question Answering*) permettent de développer de nouveaux modes interactifs pour l'accès à des collections de documents. Dans ce paradigme, l'interrogation de documents peut être vue selon deux points de vue :

- Un processus volontaire de la part de l'utilisateur qui formule délibérément une question afin d'améliorer ses connaissances antérieures sur le sujet et faciliter sa compréhension des notions abordées dans les documents. Dans ce cas le paradigme Q&A est vu comme un moteur de recherche avancé, au delà de l'interrogation par mots clés.
- Un processus implicite qui permet de créer des liens entre les documents, transformant ainsi une collection de documents en graphe de documents. Ici deux documents peuvent être mis en relation s'ils permettent de poser les mêmes questions, ou bien si les questions que soulèvent un premier document trouvent des réponses dans le second.

1. <http://archival.msh-paris.fr/>

Expériences - Le projet Archival a choisi la thématique de l'Autogestion comme support pour les travaux de recherche. Depuis les années 1960, la Bibliothèque de la Fondation Maison des Sciences de l'Homme (FMSH) a rassemblé une collection pluridisciplinaire mixte² (archives et documents) autour de l'Autogestion. Elle rassemble environ 25000 pièces : livres, revues, rapports, tracts, Parmi ces documents, nous nous intéressons tout particulièrement à la revue *Autogestion*³ (Weill, 1999) dont les numéros sont parus sur une période de 20 ans (1966-1986).

Du point de vue de traitement Automatique du Langage, les documents historiques de ce type, avec une forme de narration plus développée que des articles Wikipédia sont encore peu étudiés. Lors de travaux précédents, nous avons étudié l'adaptation de modèles de MRQA sur des données spécialisées (adaptation au domaine et transfert multilingue) (Béchet *et al.*, 2019). Nous relatons ici les premiers travaux sur un corpus d'une tout autre nature. Nous avons conduit une première série d'expériences à partir de l'annotation en question/réponse du corpus *Autogestion* avec l'objectif de vérifier dans quelle mesure un modèle à l'état de l'art appris sur de grands corpus de benchmark peut être appliqué sur un tel corpus de documents numérisés. Pour cela, nous avons utilisé un modèle de MRQA basé sur le modèle de langage pour la langue française CamemBERT (Martin *et al.*, 2020), adapté pour la tâche sur le corpus *FQUAD* (d'Hoffschmidt *et al.*, 2020) de questions/réponses en français et nous proposons une évaluation détaillée des performances en fonction du type de question posée.

Références

BÉCHET F., ALOUI C., CHARLET D., DAMNATI G., HEINECKE J., NASR A. & HERLEDAN F. (2019). Calor-quest : generating a training corpus for machine reading comprehension models from shallow semantic annotations. In *MRQA : Machine Reading for Question Answering-Workshop at EMNLP-IJCNLP 2019-2019 Conference on Empirical Methods in Natural Language Processing*.

D'HOFFSCHMIDT M., BELBLIDIA W., HEINRICH Q., BRENDLÉ T. & VIDAL M. (2020). FQuAD : French question answering dataset. In *Findings of the Association for Computational Linguistics : EMNLP 2020*, p. 1193–1208, Online : Association for Computational Linguistics. DOI : [10.18653/v1/2020.findings-emnlp.107](https://doi.org/10.18653/v1/2020.findings-emnlp.107).

MARTIN L., MULLER B., ORTIZ SUÁREZ P. J., DUPONT Y., ROMARY L., DE LA CLERGERIE É., SEDDAH D. & SAGOT B. (2020). CamemBERT : a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 7203–7219, Online : Association for Computational Linguistics.

REHM G., LEE M., MORENO-SCHNEIDER J. & BOURGONJE P. (2019). Curation technologies for cultural heritage archives : Analysing and transforming a heterogeneous data set into an interactive curation workbench. In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*, DATeCH2019, p. 117–122, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/3322905.3322909](https://doi.org/10.1145/3322905.3322909).

WEILL C. (1999). La revue autogestion comme observatoire des mouvements d'émancipation. *L'Homme et la Société*, **132**(2), 29–36.

2. Cette collection a reçu le label *Collection d'Excellence* du réseau Collex-Persée, <https://www.collexpersee.eu/le-reseau/>

3. <https://www.persee.fr/collection/autog>

Does Structure Matter? Leveraging Data-to-Text Generation for Answering Complex Information Needs

JAI2021 : Journée "Accès interactif à l'information" (AFIA-THL / ARIA / GDR TAL)

Hanane Djeddal*, Thomas Gerald*, Laure Soulier*,
Karen Pinel-Sauvagnat**, Lynda Tamine**

* Sorbonne Université, CNRS, LIP6, F-75005 Paris, France

** Université Paul Sabatier, IRIT, Toulouse, France

{hanane.djeddal, thomas.gerald, laure.soulier}@lip6.fr
{sauvagnat, tamine}@irit.fr

RÉSUMÉ

The perspective of new IR systems (e.g., search-oriented conversational systems or systems supporting complex search tasks) has fostered research on theoretical IR models either leveraging or supporting users' interactions, for instance, through question clarification ([Zamani et al., 2020](#)) or interactive ranking models ([Chen et al., 2020](#)). However, very few works focus on the way of interacting with the user in natural language, which is critical for instance for conversational systems ([Culpepper et al., 2018](#)). In this work, we focus on the upstream part of the search process, once relevant documents have been identified in response to a complex information need. Our motivation is to provide a complete and structured answer in natural language to the user on top of the retrieval model. In a close line of research, open-domain QA attempts to retrieve and reason over multiple seed passages either to extract ([Asai et al., 2020](#); [Chen et al., 2017](#)) or to generate in a natural language form ([Song et al., 2018](#); [Wang et al., 2018](#); [Tan et al., 2018](#)) answers to open-domain questions. While most QA problems target a single-span answer ([Yang et al., 2018](#)) included in one document, open-ended queries are characterized by multiple facets ([White & Roth, 2009](#); [Wildemuth & Freund, 2012](#)) that could target a multiple-span answer.

Our objective is to generate an answer that covers the multiple facets of an open-ended query using as input, an initial ranked list of documents. We, therefore, envision solving complex information needs with generative models (seq-to-seq models), particularly from the perspective of data-to-text generation. This last category of models puts its attention on the notion of structure which we believe is important when answering multi-faceted queries. One particular approach retained our attention : a content selection and planning pipeline ([Puduppully et al., 2019](#)) which aims at structuring the answer by generating intermediate plans. This intermediary step reinforces the factualness and the coverage of the generated text since : 1) it organizes the data structure in a latent form to better fit with the generated output, and 2) it provides a structure to the generated answer based on the elements of the initial data. The designed model is driven by the intuition that the response should be guided by a structure to cover most of the query facets. This prior is modeled through a hierarchical plan which corresponds to a textual object relating the structure of the response with multiple-level titles (titles, subtitles, etc).

The experimental evaluation is performed using the TREC Complex Answer Retrieval (CAR) dataset ([Dietz et al., 2017](#)). We evaluate both the generated answer and its corresponding structure and show

the effectiveness of planning-based models in comparison to a text-to-text model. The results confirm our intuition that data-to-text generation models can be leveraged to generate in a single-turn, a natural language and structured answer, providing thus, an alternative to the interactive updating of queries and document rankings as answer. We believe that our work opens up novel areas of investigation including answer generation and explanation in conversational systems for IR.

Références

- ASAI A., HASHIMOTO K., HAJISHIRZI H., SOCHER R. & XIONG C. (2020). Learning to retrieve reasoning paths over wikipedia graph for question answering. In *ICLR*.
- CHEN D., FISCH A., WESTON J. & BORDES A. (2017). Reading wikipedia to answer open-domain questions. In *Proceedings of the 55th Annual Meeting of ACL 2017*, p. 1870–1879 : ACL.
- CHEN L., TANG Z. & YANG G. H. (2020). Balancing reinforcement learning training experiences in interactive information retrieval. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020*, p. 1525–1528.
- CULPEPPER J. S., DIAZ F. & SMUCKER M. D. (2018). Research frontiers in information retrieval : Report from the third strategic workshop on information retrieval in lorne (swirl 2018). *SIGIR Forum*, **52**(1), 34–90.
- DIETZ L., VERMA M., RADLINSKI F. & CRASWELL N. (2017). TREC complex answer retrieval overview. In *TREC : (NIST)*.
- PUDUPPULLY R., DONG L. & LAPATA M. (2019). Data-to-text generation with content selection and planning. p. 6908–6915 : The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019.
- SONG L., WANG Z., HAMZA W., ZHANG Y. & GILDEA D. (2018). Leveraging context information for natural question generation. In M. A. WALKER, H. JI & A. STENT, Éd., *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, p. 569–574 : Association for Computational Linguistics.
- TAN C., WEI F., YANG N., DU B., LV W. & ZHOU M. (2018). S-net : From answer extraction to answer synthesis for machine reading comprehension. In *AAAI*.
- WANG Y., LIU K., LIU J., HE W., LYU Y., WU H., LI S. & WANG H. (2018). Multi-passage machine reading comprehension with cross-passage answer verification. *CoRR*.
- WHITE R. W. & ROTH R. A. (2009). *Exploratory Search : Beyond the Query-Response Paradigm*. [San Rafael, Calif.] : Morgan & Claypool Publishers.
- WILDEMUTH B. M. & FREUND L. (2012). Assigning search tasks designed to elicit exploratory search behaviors. *HCIR '12 : Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval ACM*.
- YANG Z., QI P., ZHANG S., BENGIO Y., COHEN W., SALAKHUTDINOV R. & MANNING C. D. (2018). HotpotQA : A dataset for diverse, explainable multi-hop question answering. p. 2369–2380 : Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing ACL.
- ZAMANI H., DUMAIS S. T., CRASWELL N., BENNETT P. N. & LUECK G. (2020). Generating clarifying questions for information retrieval. In *WWW '20 : The Web Conference 2020*, p. 418–428.

Ero.Dot, un agent conversationnel pour le théâtre par combinaison de connaissances discrètes et continues

jaii2021 : Journée "Accès interactif à l'information"
(AFIA-THL / ARIA / GDR TAL)

Aniss Bentebib, Jean-Pierre Chevallet, Didier Schwab
Université Grenoble Alpes

`aniss.bentebib/jean-pierre.chevallet/didier.schwab@univ-grenoble-alpes.fr`

1 Introduction

"Terres rares"¹ est un cyber-opéra qui raconte l'histoire d'une expédition géologique associant humains et robots faisant fiasco aux confins du grand Nord et se soldant par un carnage. Mandatés par les commanditaires de l'expédition, deux enquêteurs cherchent à établir les faits. Encore opérationnel, le cyborg Ero.dot (faisant parti de l'expédition) révèle les tenants du drame avant de générer une ode au Vivant.

Le système que nous avons développé a pour objectif de donner vie numériquement à Ero.Dot via un avatar dans une application mobile. En effet, après la pièce, le spectateur est amené à poursuivre certains pans narratifs non explorés durant le spectacle. Pour cela, il interagit par le dialogue avec un agent conversationnel en lui posant des questions et en essayant de découvrir plus en détail les circonstances du drame. Pour lui donner substances, il a été décidé d'associer deux formats de connaissances : les connaissances discrètes et continues. Dans la suite de cet article nous détaillerons ce que nous entendons par ces termes et l'intérêt de cette démarche.

2 Définitions

Les connaissances "discrètes" sont ce qu'on pourrait appeler des connaissances sécables et identifiables de façon claires, ordonnées et dissociables l'une de l'autre dans un espace. Ex : Les graphes de connaissances (Hogan *et al.*, 2021) dont les sommets sont des mots et les arcs des relations. Elles ont pour principaux intérêts d'être intelligibles et facilement manipulables en langage naturel. D'un autre côté, nous avons des connaissances "continues". Nous appelons ces connaissances ainsi car elles sont plongées dans des espaces continus, dont les zones sont difficilement sécables. Ex : Les plongements de mots (contextuels (Devlin *et al.*, 2019) ou non). L'intérêt majeur réside dans leurs propriétés calculatoires. En effet, ces connaissances sont bien souvent représentées par des vecteurs qui permettent l'usage rapide de ces derniers dans des algorithmes de machine learning par exemple.

1. <https://www.atelier-arts-sciences.eu/Terres-rares>

3 Motivations

La combinaison efficace de ces deux types de connaissances pourrait permettre de proposer de nouvelles façons d'interagir avec des agents conversationnels. En effet, ils seraient bien plus maléables, personnalisables, plus "intelligents" au regard des connaissances stockées et donc plus efficaces pour réaliser certaines tâches. Ex : Enseigner une langue, chercher des recommandations de livres en échangeant quelques mots avec un utilisateur (Gao *et al.*, 2021), etc. De plus, le dialogue est un moyen assez naturel pour les humains de chercher une information. Il est possible que parfois, de simples mots clefs ne suffisent pas à exprimer pleinement notre désir d'information (Boussaha *et al.*, 2019). Parce que ce que nous cherchons est trop complexe ou bien très spécifique. Que faisons-nous lorsque nous avons des difficultés à exprimer quelque chose à quelqu'un ? Nous paraphrasons, échangeons et répondons à des requêtes de désambiguations de notre interlocuteur. C'est pourquoi nous pensons les systèmes de recherche d'information peuvent gagner en précision en utilisant de nouveaux canaux comme le dialogue (Gao *et al.*, 2020). Pour terminer, dans le cadre du projet de cyber-opéra, il était question de permettre à des artistes de définir et de créer des agents conversationnels facilement. La combinaison de différentes formes de connaissances permet cela grâce à l'efficacité du format continu et l'intelligibilité des connaissances discrètes dont le format est généré à partir du langage naturel.

4 Mise en œuvre

Notre première approche se base sur l'agent conversationnel Blender (Roller *et al.*, 2020). C'est un modèle génératif de dialogues qui introduit un mécanisme de recherche d'information. Ce mécanisme consiste à prendre un contexte/historique de conversation puis un ensemble de candidats à la succession de la conversation et de leur attribuer un score de pertinence durant la phase d'apprentissage. Ce mécanisme est assuré par un modèle que les auteurs ont introduit précédemment (Humeau *et al.*, 2020). Le système Blender est en anglais c'est pourquoi pour avoir un système en français nous avons un traducteur des entrées utilisateurs vers l'anglais et en sortie un traducteur anglais vers français.

Le générateur est de type séquence vers séquence (*Seq2Seq*) basé sur l'architecture transformers (Vaswani *et al.*, 2017) entraîné avec 2.7 milliards de paramètres (modèle blender_3B - 2,7B)². L'idée sous-jacente derrière ce système est d'éviter au modèle de générer des réponses trop redondantes et trop ternes. En allant chercher de l'information supplémentaire (des pages en langage naturel) dans wikipédia si le contexte est trop pauvre, on permet à l'agent de s'exprimer avec plus de contexte et de diversité. Pour se faire, un classifieur binaire est entraîné à prédire si oui ou non la requête en cours nécessite d'être augmentée. Si oui, un index inversé basé sur TF-IDF est utilisé sur un dump de wikipédia pour conditionner la réponse (Roller *et al.*, 2020).

4.1 Apprentissage

Dans la Fig.1, le processus se décompose de la sorte : Le créateur d'agent définit des traits de personnalités et/ou des connaissances que l'agent conversationnel doit être capable de restituer dans

2. <https://parl.ai/projects/recipes/>

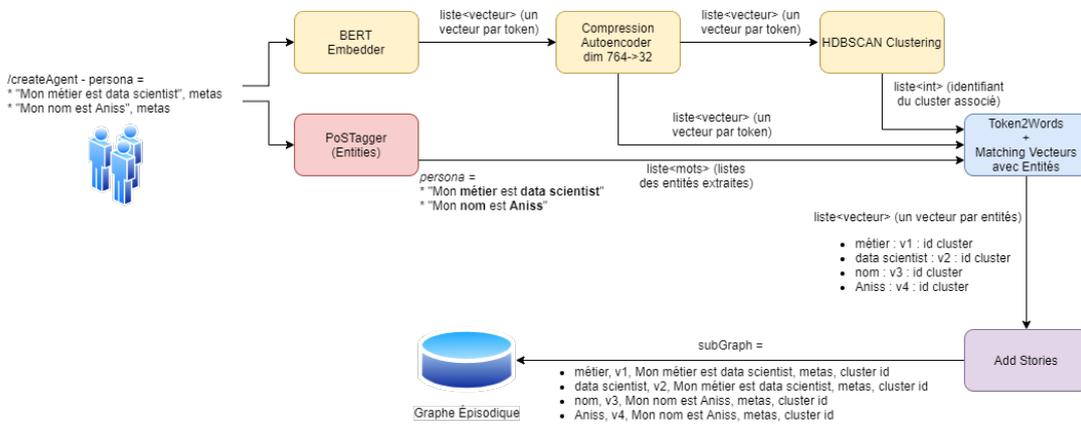


FIGURE 1 – Schéma d’exécution des différents processus d’apprentissage pour la création d’un agent

la conversation. On appelle ces phrases des personas ou des souvenirs. Un persona ou un souvenir est une phrase qui décrit une caractéristique propre à la personnalité de l’agent conversationnel, par exemple : "Je m’appelle Ero.DOT" serait un persona. Ces phrases sont plongées contextuellement dans un espace vectoriel grâce à BERT Base. Ces vecteurs de mots sont compressés au moyen d’un autoencoder pour éviter le problème de la malédiction des hautes dimensions (Pestov, 2013), ils sont ensuite envoyé dans un graphe de densité HDBSCAN (Malzer & Baum, 2020) permettant à l’inférence d’accéder aux connaissances de manières plus efficaces. Pour résumer, les mots des souvenirs sont encodés contextuellement, réduit en terme de dimension puis on crée une carte de densité de l’espace dans lequel ils sont plongés pour permettre de récupérer le voisinage efficacement durant l’inférence. Une fois ce processus exécuté on envoie les souvenirs dans un graphe de connaissance épisodique (on stocke le timestamp pour une recherche d’information temporelle).

4.2 Inférence

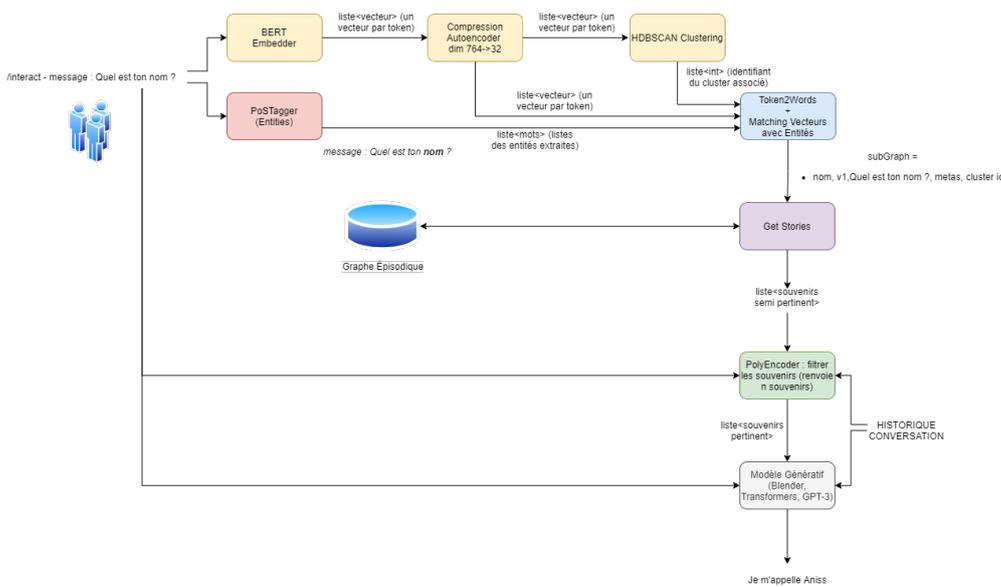


FIGURE 2 – Schéma d’exécution des différents processus régissant l’interaction entre un utilisateur et son agent

En Fig.2 est décrit le processus d’inférence de notre système. Pour se faire, les requêtes de l’utilisateur

suivent le même processing que durant la phase d'apprentissage. Pour chaque mot de la requête, on les plonge contextuellement, on les réduit puis, grâce au modèle HDBSCAN, on récupère les mots voisins aux différents mots de la requête. Un voisinage proche signifie que les mots appartenant à ce dernier ont une proximité sémantique contextuelle importante donc sont susceptibles d'être associés à des souvenirs comportant l'information nécessaire à l'élaboration d'une réponse cohérente par le modèle génératif. Nous obtenons une liste de souvenirs qui subissent un second filtre (le premier étant la proximité dans l'espace de densité) qui consiste à classer les souvenirs extraits de HDBSCAN en prenant cette fois en compte l'historique de la conversation. Après ce filtre nous gardons seulement $\min(\#\text{souvenirs}, n)$ avec n étant un hyperparamètre que nous avons fixé à 5 empiriquement (grid search). Pour terminer le modèle génératif prend en entrée l'historique, les n souvenirs les plus pertinents et la requête de l'utilisateur pour générer une réponse.

5 Conclusion et perspectives

Ce système présente l'avantage de permettre la création d'agents conversationnels avec des personnalités complexes, sous-entendu plus que 5 lignes de personnalités imposées dans le papier de Blender. Il présente également la possibilité de stocker de l'information et de la restituer dans une conversation dans un processus tout à fait intelligible et rend tout à fait possible la création d'agents conversationnels ayant des personnalités très particulières comme celle d'Ero.Dot. Dans un second temps nous souhaitons mettre en place un pipeline d'évaluation pour permettre de comparer notre approche avec des approches similaires. Techniquement parlant il y a plusieurs choses améliorables notamment autour de la complexité du modèle d'un point de vue implémentation. Nous réfléchissons également à d'autres façon de combiner différents formats de connaissances, nous allons essayer les réseaux de neurones graphiques ainsi que les plongements contextuels de graphes. Nous avons également pour projet de créer un ensemble de donnée pour le dialogue en français permettant l'entraînement de systèmes à grande échelle. L'objectif à terme est de proposer un système d'agent conversationnel en français sans passer par des modules de traduction.

6 Remerciements

Ce travail a été réalisé en collaboration avec le CEA (le Commissariat à l'Énergie Atomique et aux Énergies Alternatives) qui a fondé "L'atelier Art Science" et l'entreprise Hoomano qui finance et soutient le projet de cyber-opéra. Ce travail a été partiellement financé par MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

Références

BOUSSAHA B. E. A., HERNANDEZ N., JACQUIN C. & MORIN E. (2019). Deep retrieval-based dialogue systems : A short review.

DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2019). Bert : Pre-training of deep bidirectional transformers for language understanding.

- GAO C., LEI W., HE X., DE RIJKE M. & CHUA T.-S. (2021). Advances and challenges in conversational recommender systems : A survey. *AI Open*, **2**, 100–126. DOI : <https://doi.org/10.1016/j.aiopen.2021.06.002>.
- GAO J., XIONG C. & BENNETT P. (2020). Recent advances in conversational information retrieval. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, p. 2421–2424, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/3397271.3401418](https://doi.org/10.1145/3397271.3401418).
- HOGAN A., BLOMQUIST E., COCHEZ M., D'AMATO C., MELO G. D., GUTIERREZ C., KIRRANE S., GAYO J. E. L., NAVIGLI R., NEUMAIER S. & ET AL. (2021). Knowledge graphs. *ACM Computing Surveys*, **54**(4), 1–37. DOI : [10.1145/3447772](https://doi.org/10.1145/3447772).
- HUMEAU S., SHUSTER K., LACHAUX M.-A. & WESTON J. (2020). Poly-encoders : Transformer architectures and pre-training strategies for fast and accurate multi-sentence scoring.
- MALZER C. & BAUM M. (2020). A hybrid approach to hierarchical density-based cluster selection. *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. DOI : [10.1109/mfi49285.2020.9235263](https://doi.org/10.1109/mfi49285.2020.9235263).
- PESTOV V. (2013). Is the k-nn classifier in high dimensions affected by the curse of dimensionality? *Computers Mathematics with Applications*, **65**(10), 1427–1437. Grasping Complexity, DOI : <https://doi.org/10.1016/j.camwa.2012.09.011>.
- ROLLER S., DINAN E., GOYAL N., JU D., WILLIAMSON M., LIU Y., XU J., OTT M., SHUSTER K., SMITH E. M., BOUREAU Y.-L. & WESTON J. (2020). Recipes for building an open-domain chatbot.
- VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER L. & POLOSUKHIN I. (2017). Attention is all you need.

Evaluation des systèmes de recherche d'information interactifs : exemple du Bouclage de Pertinence implicite par suivi oculaire

Philippe Mulhem, Francis Jambon, Lucas Albarede
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
{Philippe.Mulhem, Francis.Jambon}@imag.fr,
lucas.albarede@etu.univ-grenoble-alpes.fr

1 Introduction

Un système de recherche d'information (SRI) est par définition un système interactif qui vise à fournir à un utilisateur les documents pertinents pour son besoin d'information. L'évaluation des SRI est un élément fondamental de ce domaine de recherche. Historiquement, l'aspect interactif a été peu intégré dans l'évaluation des SRI : il rentre en conflit avec le fait que l'on souhaite évaluer les systèmes de manière reproductible et "en grand", c'est-à-dire avec de nombreuses requêtes portant sur de nombreux documents, et qu'il n'est souvent pas possible, d'un simple point de vue logistique, de réaliser cela avec les humains. Nous proposons une approche basée sur la simulation d'un ensemble d'utilisateurs afin de se confronter à ce problème, en prenant comme exemple le cas du bouclage de pertinence avec du suivi oculaire, pour la recherche de pages web avec utilisation de snippets.

2 Exemple du bouclage de pertinence implicite par suivi oculaire

Le bouclage de pertinence implicite utilisant le suivi oculaire s'appuie sur l'hypothèse qu'il est possible de déterminer les centres d'intérêts implicites d'un utilisateur, en étudiant la façon dont celui-ci lit les termes d'une page de résultats (Sunjeelee *et al.*, 2020). L'état de l'art sur ce sujet étant restreint, nous avons créé notre propre modèle de bouclage de pertinence qui intègre des caractéristiques obtenues à partir du suivi oculaire d'un utilisateur. Ce modèle, appelé modèle Bo1ET (Albarede, 2019), est une extension du modèle Bo1 (Amati & Van Rijsbergen, 2002) de pseudo-bouclage de pertinence, que nous avons adapté pour la recherche sur le Web (SERP), et intégrant des caractéristiques "classiques" (Holmqvist *et al.*, 2011) du suivi oculaire, comme le temps de lecture moyen d'un mot. Le paramétrage puis l'évaluation "en grand" d'un tel système, tout en intégrant des éléments provenant d'une interaction avec l'utilisateur reproductible, nécessiterait d'avoir recours à plusieurs centaines d'utilisateurs (afin de minimiser la variabilité inter-utilisateurs), sur l'ensemble d'une collection (afin d'être comparable avec d'autres travaux), et dans des locaux dédiés (afin de pouvoir utiliser le dispositif de suivi oculaire), ce qui n'est pas réalisable en pratique. C'est pourquoi, nous nous sommes tournés vers l'utilisation de la simulation du comportement des utilisateurs.

3 Cadre général de l'évaluation par simulation

Quand on parle de simulation de suivi oculaire, la question de quoi simuler, et pour quel objectif, se pose. Dans notre cas, nous avons choisi de simuler non pas directement les parcours oculaires, mais les caractéristiques de haut niveau : notre argument principal de ce choix est que les simulations des parcours oculaires sont d'une très grande complexité (Yarbus, 1967), et pas nécessaire dans notre cas. Notre système reposant sur des caractéristiques générées à partir des parcours oculaires,

nous choisissons de privilégier la simulation au niveau de ces caractéristiques : dans ce cas, nous ne simulons que les éléments directement liés à notre problématique. Le désavantage d’une telle solution est qu’il n’est possible d’évaluer que les systèmes interactifs utilisant exclusivement ces caractéristiques. Le choix de ces caractéristiques est donc d’un intérêt primordial pour l’acceptation de nos propositions. De manière plus formelle, nous appelons :

- caractéristiques hors contexte applicatif, F_{hc} , les suivis oculaires (saccades et fixations). Ces éléments ont classiquement acquis par les eye-trackers quelque soit le contexte applicatif considéré ;
- caractéristiques en contexte applicatif, F_{ec} , les caractéristiques construites à partir des F_{ec} qui décrivent des éléments du domaine applicatif. Dans notre cas, ces caractéristiques sont celles utilisées par un SRI interactif, par exemple, le temps de lecture d’un mot pertinent, le temps de lecture d’une snippet, etc.

Le passage des F_{hc} au F_{ec} est donc fondamental et doit être défini avec précision.

4 Adaptation du cadre général à l’évaluation de la RF

De manière à pouvoir comparer notre approche à d’autres travaux, et à la maintenir dans le paradigme de Cranfield, nous utilisons les documents du corpus, les requêtes et les jugements de pertinence issus de collections de test classiques. Nous avons étendu ces données pour l’évaluation interactive en y ajoutant des indications sur la pertinence des mots présentés dans les SERP, et la pertinence des snippets (déduite directement de celle des documents). Dans notre exemple, chaque caractéristique de contexte applicatif F_{ec} (ou feature) est associée à une probabilité de valeur pour chaque mot de chaque snippet présenté. Cette feature est calculée à partir des caractéristiques associées au mot qui dépendent de la pertinence du mot (3 valeurs de pertinences : pertinent/non-pertinent/topical), et des caractéristiques associées à la snippet (2 valeurs de pertinences : pertinent/non-pertinent). Notons que les valeurs de pertinences sont connues du générateur de features, qui simule l’humain, mais pas du SRI. Nous restons donc cohérent avec le fait que le SRI réagit uniquement aux actions simulées, actions qui elles dépendent de la pertinence des éléments présentés à l’interface, mais le SRI n’a pas connaissance des valeurs de pertinence. Dans notre exemple, nous avons défini par un jugement d’experts la pertinence des mots des snippets, en se basant sur les requêtes initiales. La pertinence d’un snippet est déduite de la pertinence des documents définie dans la collection de test (données existantes dans les collections de test actuelles)(Mizzaro, 1997).

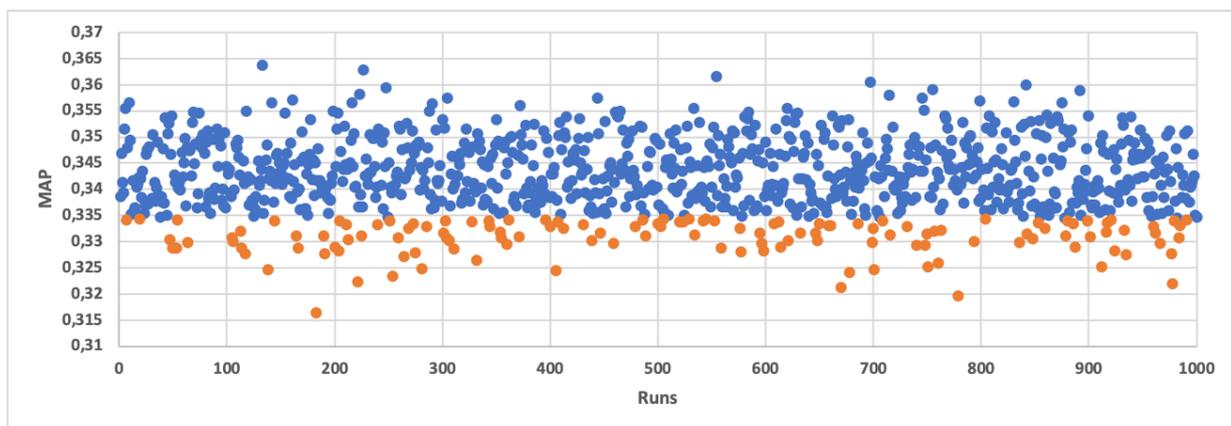


FIGURE 1 – MAP de Bo1EG pour 1000 utilisateurs simulés ; axe des X : numéro d’utilisateur ; axe des Y : MAP de Bo1EG comparée à Bo1mod (bleu : Bo1EG est supérieur à Bo1mod ; rouge : Bo1EG est inférieur à Bo1mod).

5 Mise en œuvre

Une première étude a été effectuée sur 8 requêtes sur les corpus TREC-5 et TREC-6, avec les requêtes de TREC-6 ad-hoc (Voorhees & Harman, 2000). Dans un premier temps, la feature utilisée a été mesurée sur un ensemble restreint d'utilisateurs de manière à en déterminer sa gaussienne. Elle a ensuite été simulée à partir de cette dernière. La figure 1 présente les résultats sur 1000 simulations qui montrent un avantage (en MAP) que Bo1ET améliore dans 867 cas (points bleus) les résultats de Bo1 classique, et dégrade dans 133 cas (points rouges).

6 Conclusion et Perspectives

Ces travaux préliminaires montrent qu'il est envisageable, tout en restant dans le cadre du paradigme de Cranfield, d'évaluer un système de recherche d'information interactif via la simulation des utilisateurs. Nous travaillons actuellement à généraliser et outiller cette approche afin d'étendre son domaine de validité et d'utilisation.

7 Remerciements

Ces travaux ont été en partie financés par le Laboratoire d'Informatique de Grenoble (UMR 5217) via les projets "Emergence" GELATI et IIRSuts.

Références

- ALBAREDE L. (2019). *Relevance Feedback Enhanced with Eye Movements Analysis*. Rapport interne, Université Grenoble Alpes, Master of Science in Informatics at Grenoble.
- AMATI G. & VAN RIJSBERGEN C. J. (2002). Probabilistic models of information retrieval based on measuring the divergence from randomness. *ACM Trans. Inf. Syst.*, **20**(4), 357–389.
- HOLMQVIST K., NYSTRÖM M., ANDERSSON R., DEWHURST R., HALSZKA J. & VAN DE WEIJER J. (2011). *Eye Tracking : A Comprehensive Guide to Methods and Measures*. United Kingdom : Oxford University Press.
- MIZZARO S. (1997). Relevance : The whole story. *Journal of the American Society for Information Science*, (48), 810–832.
- SUNGEELEE V., JAMBON F. & MULHEM P. (2020). Proof of Concept and Evaluation of Eye Gaze Enhanced Relevance Feedback in Ecological Context. In *Proceedings of the Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020), Samatan, Gers, France, July 6-9, 2020*, volume 2621 de *CEUR Workshop Proceedings* : CEUR-WS.org.
- VOORHEES E. M. & HARMAN D. (2000). Overview of the sixth text retrieval conference (trec-6). *Inf. Process. Manage.*, **36**(1), 3–35. DOI : [10.1016/S0306-4573\(99\)00043-6](https://doi.org/10.1016/S0306-4573(99)00043-6).
- YARBUS A. L. (1967). *Eye Movements and Vision*. Plenum. New York.

Interactive Query Clarification and Refinement via User Simulation

Pierre Erabcher, Ludovic Denoyer*, Laure Soulier

Sorbonne Universite, Currently working at Ubisoft*

`pierre.erbacher@isir.upmc.fr`

Keywords : Interactive Information Retrieval, Query Clarification, User Simulation

Understanding information need is a long-standing issue in Information Retrieval (IR) (Cronen-Townsend & Croft, 2002; Jansen *et al.*, 2000; Sanderson, 2008), often highlighted by the difficulty for users to formulate open-ended information needs into queries. Queries are thus often under-specified and/or ambiguous (Jansen *et al.*, 2000). Depending on the user or the context, the same query may refer to different intents. For instance, the query "Orange" may refer to several topics, including company names, locations, songs titles, ... To tackle this issue, numerous works have been proposed.

A promising approach has been proposed in (Aliannejadi *et al.*, 2019) to clarify information needs by proactively interacting with the user. The classic workflow for asking clarifications is based on three main steps (Aliannejadi *et al.*, 2019) : 1) the IR system produces a clarifying question for the user, 2) the latter provides an answer or selects an option, and 3) the IR system ranks documents according to the user's feedback. The pioneering work (Aliannejadi *et al.*, 2019) aims at generating clarifying questions by 1) retrieving a predefined set of questions using a Bert-based model and 2) at each turn, selecting the best query through a conversation history-driven model. The user's answer corresponds to a predefined text built using crowd-sourcing. One drawback of this approach is that the multi-turn conversation is log-based, interactively simulated using predefined logs of conversation history (i.e., sequence of questions/answers obtained by HITS). This simulated conversation defined a priori without interaction with the proposed question selection model might hinder the evaluation performance in the sense that we are not sure about the soundness of the conversation flow. Other work (Zamani *et al.*, 2020; Wu *et al.*, 2018; Guo *et al.*, 2011; Santos *et al.*, 2012; Ozertem *et al.*, 2012) tackle this issue by proposing generative models, that create clarification question or query suggestions. But they do not address the multi-turn framework, stopping the clarification process at the first interaction.

In this paper, we propose to build a fully simulated query clarification framework allowing multi-turn interactions between IR and user agents. Following (Aliannejadi *et al.*, 2019), the IR agent identifies candidate queries and ranks them in the context of the user-system interactions to clarify the initial query issued by the user agent. We particularly target simple information needs, multiple information needs are let for future work since it might impact the modeling of the query ranking function. Our framework can be seen as a basis and a proof-of-concept for future work willing to integrate sequential models (namely reinforcement learning models) for question clarification. It is worth noting that large language models relying on attention mechanisms (transformers) are not yet well suited to handle sequential interactions and long-term planning, as current models are hardly trainable with current reinforcement learning algorithms (Chen *et al.*, 2020). Therefore, all agent components in our framework are based on continuous and simple models. To validate our simulation framework, we conduct an experimental analysis on the MS Marco dataset. We show the benefit of multi-turn

interactivity and evaluate the effectiveness of different question selection strategies.

In what follows, we present the different components behind the IR system and user agents.

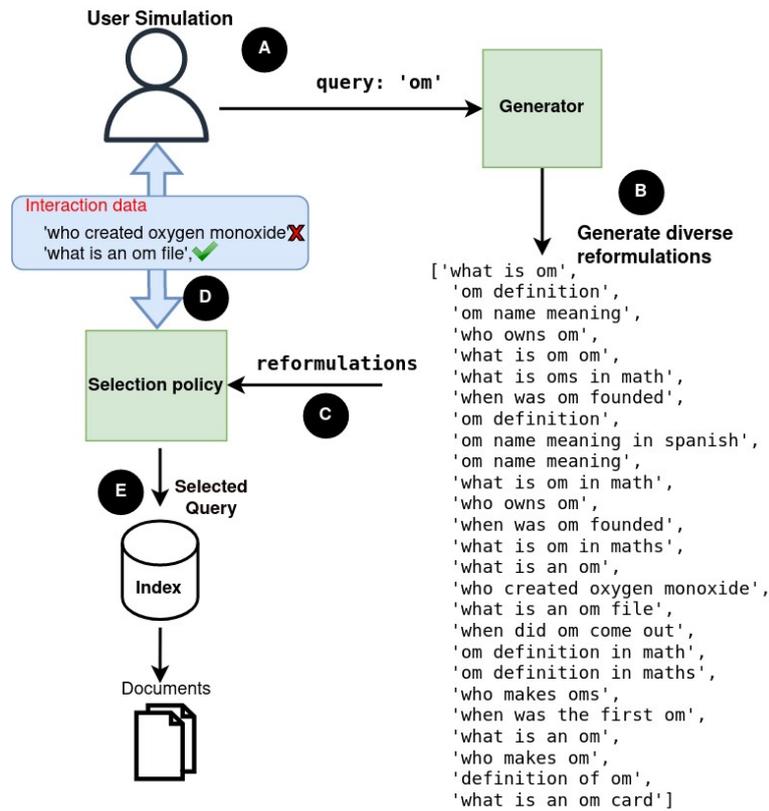


FIGURE 1 – Query clarification simulation framework

A Language model generates a fixed set of reformulation (Generator) given a user query. At each time step the selection policy evaluates a new score for each reformulation given previous user feedback. It selects 2 reformulations following a strategy. The user simulation selectd the preferred reformulation. After T interactions, the best ranked reformulation is used to retrieve documents. The experiment is performed on msmarco with various selection strategy.

Références

- ALIANNEJADI M., ZAMANI H., CRESTANI F. & CROFT W. B. (2019). Asking clarifying questions in open-domain information-seeking conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'19*, p. 475–484, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/3331184.3331265](https://doi.org/10.1145/3331184.3331265).
- CAMPILLOS LLANOS L., THOMAS C., BILINSKI E., ZWEIGENBAUM P. & ROSSET S. (2019). Designing a virtual patient dialogue system based on terminology-rich resources : Challenges and evaluation. *Natural Language Engineering*, p. 1–38. DOI : [10.1017/S1351324919000329](https://doi.org/10.1017/S1351324919000329).
- CHEN J. Z., YU S. Y. & WANG H. (2020). Exploring fluent query reformulations with text-to-text transformers and reinforcement learning. *ArXiv*, **abs/2012.10033**.
- CRONEN-TOWNSEND S. & CROFT W. B. (2002). Quantifying query ambiguity. In *Proceedings of the Second International Conference on Human Language Technology Research, HLT '02*, p. 104–109, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.
- GUO J., CHENG X., XU G. & ZHU X. (2011). Intent-aware query similarity. In *CIKM '11*.

- JANSEN B., SPINK A. & SARACEVIC T. (2000). Real life, real users, and real needs : A study and analysis of user queries on the web. *Information Processing and Management*, **36**(2), 207–227. DOI : [10.1016/S0306-4573\(99\)00056-4](https://doi.org/10.1016/S0306-4573(99)00056-4).
- LLANOS L. C., GROUIN C., LOUËT A. L.-L. & ZWEIGENBAUM P. (2019). Initial experiments for pharmacovigilance analysis in social media using summaries of product characteristics. In *MEDINFO 2019*, volume 264 de *Studies in Health Technology and Informatics*, p. 60–64, Lyon, France : Amsterdam : IOS Press.
- OZERTEM U., CHAPELLE O., DONMEZ P. & VELIPASAOGLU E. (2012). Learning to suggest : A machine learning framework for ranking query suggestions. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, p. 25–34, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/2348283.2348290](https://doi.org/10.1145/2348283.2348290).
- SANDERSON M. (2008). Ambiguous queries : Test collections need more sense. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, p. 499–506, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1390334.1390420](https://doi.org/10.1145/1390334.1390420).
- SANTOS R. L. T., MACDONALD C. & OUNIS I. (2012). Learning to rank query suggestions for adhoc and diversity search. *Information Retrieval*, **16**, 429–451.
- WU B., XIONG C., SUN M. & LIU Z. (2018). Query suggestion with feedback memory network. *Proceedings of the 2018 World Wide Web Conference*.
- ZAMANI H., DUMAIS S., CRASWELL N., BENNETT P. & LUECK G. (2020). *Generating Clarifying Questions for Information Retrieval*, In *Proceedings of The Web Conference 2020*, p. 418–428. Association for Computing Machinery : New York, NY, USA.

projets TactiNET et TagThunder : quelle modalité pour un *skimming* et un *scanning* non visuels de l'information textuelle ?

Fabrice Maurel

Universite de Caen Normandie
fabrice.maurel@unicaen.fr

RÉSUMÉ

La synthèse vocale et la plage braille sont les principales technologies utilisées par les déficients visuels. Cependant, elles restent à la fois peu pratiques, peu discrètes et peu efficaces sur les supports mobiles pour lesquels l'interaction tactile est indispensable (Safi *et al.*, 2019).



La figure ci-contre illustre la difficulté d'appréhender non visuellement une page Web. Si la page de gauche permet en quelques secondes de prélever un grand nombre d'informations, celle de droite n'offre pas de telles possibilités. Il s'agit pourtant de la même page selon qu'elle est produite par un outil de visualisation ou destinée à des sorties tactilo-orales. La difficulté est de nature différente sur les dispositifs tactiles. Les solutions logicielles d'accessibilité nativement embarquées dans les tablettes et les smartphones modifient

moins fortement la structure visuelle mais synthétisent à l'oral le texte au fur et à mesure qu'il est survolé par le doigt. Cette solution est intéressante mais rébarbative dès lors qu'il s'agit de parcourir des documents nouveaux, et donc en particulier dans le cadre d'une navigation Web : le non-voyant doit préalablement percevoir, interpréter et mettre en relation les bribes de synthèse de parole produites par son interaction avec l'organisation physique et logico-thématique des éléments de la page. Pour cela il déplace son doigt sur la quasi-totalité de l'écran pour procéder à une phase d'apprentissage lourde et quelque peu aléatoire ; en tous les cas encore incomplète pour faire émerger des stratégies de lecture rapide. De fait, la plupart des utilisateurs ne le font que rarement et ont une pratique très utilitaire des dispositifs tactiles, se cantonnant aux fonctionnalités des sites et des interfaces qu'ils connaissent parfaitement. Dans un objectif de réduction de la fracture numérique, il est impératif de permettre une appréhension non visuelle qui soit à la fois globale et naturellement interactive, en augmentant les capacités perceptives des non-voyants par un accès à la structure informationnelle et organisationnelle des pages Web. Il est largement admis que les propriétés visuelles du texte agissent sur de nombreuses dimensions : (1) elles jouent sur la lisibilité des documents et donc sur leur accessibilité cognitive ; (2) à l'instar de la prosodie, elles véhiculent une part de la sémantique du message ; (3) par la qualité d'affordance qu'elles procurent au document, elles exploitent nos tendances perceptives naturelles pour suggérer des parcours interprétatifs cohérents ; (4) elles génèrent des possibilités nouvelles propres aux intentions de lecture individuelles. En cela, elles soutiennent l'émergence de stratégies créatives par expérience ou par sérendipité ; (5) Elles favorisent la capacité d'action de l'œil à combiner rapidement des opérations de prélèvement d'information à la fois locales et globales. C'est cette interaction et cette dynamique, qui sous-tend toutes les autres, que nous souhaitons nous employer à conserver lors de la transposition des propriétés visuelles dans de nouvelles modalités sensorielles. Nous nous intéressons à deux stratégies procédant d'un enchaînement de processus

mentaux fréquents lors de la lecture silencieuse de documents par un voyant : le lecteur jette un premier regard sur tout ou partie de la page et en effectue un survol quasi-instantané (*skimming*) ; il initie ensuite une recherche rapide d'indices visuels et langagiers, sélectionnés en fonction des intentions de lecture (*scanning*) ; ces deux stratégies, plus ou moins conscientisées, peuvent se répéter selon différentes combinaisons jusqu'à la satisfaction d'objectifs individuels. La mise page et la typographie prennent une part déterminante dans le succès et l'efficacité de ces processus. Notre réflexion porte sur la possibilité de les rendre accessibles aux non-voyants ; autrement dit, comment favoriser le développement de stratégies de *skimming* et de *scanning* non visuels en s'appuyant sur une transposition orale et tactile de la structure visuelle des documents ? Une critique essentielle que nous formulons à l'endroit de nombreuses solutions de la littérature porte sur l'approche méthodologique qui est posée. Elles s'attachent, pour la plupart, à réduire la charge cognitive en recherchant dans la page Web les informations « pertinentes » et en éliminant les perturbations induites par les éléments « périphériques » ; ou encore en intégrant des techniques de résumés. De manière générale il s'agit de simplifier le contenu pour le rendre plus digeste aux modalités tactiles ou orales. Ce faisant, le concepteur considère que l'amélioration de l'accessibilité doit sacrifier une certaine richesse de contenu. Une définition de l'accessibilité universelle affirme qu'elle est le caractère d'un produit, procédé, service, environnement ou de l'information qui, dans un but d'équité et dans une approche inclusive, permet à toute personne de réaliser des activités de façon autonome et d'obtenir des résultats équivalents. C'est cette récente prise en compte de la capacité d'autodétermination de l'utilisateur d'un système interactif qui nous semble majeure et prioritaire à évaluer. Notre ambition est de ne pas nous identifier subjectivement à un lecteur théorique dont nous connaîtrions les intentions, mais plutôt de fournir à l'utilisateur réel l'ensemble de l'information visuelle dans sa complexité ; à notre charge de trouver les stimuli adéquats pour favoriser l'interaction et le déroulement de la boucle perception/action ; à la charge de chaque utilisateur d'apprendre à les maîtriser, à se les approprier, à pérenniser ses propres interprétations ; à la charge du temps et de la pratique de faire émerger des stratégies nouvelles de prélèvement d'information. Nous présenterons deux projets, TactiNET (Maurel *et al.*, 2020) et TagThunder (Ramesh *et al.*, 2021), qui s'inscrivent dans cette démarche et pour lesquels nous décrirons une approche dirigée par la tâche ainsi que les premiers résultats ; ces derniers s'appuient sur des algorithmes de partitionnement multicritères suivant l'hypothèse que les unités élémentaires d'une page Web peuvent être comparées simultanément sur une triple base : contenu textuel, forme visuelle et relations logiques.

Références

MAUREL F., DIAS G., SAFI W., ROUTOURE J.-M. & BEUST P. (2020). Layout transposition for non-visual navigation of web pages by tactile feedback on mobile devices. *Micromachines*, **11**(4). DOI : [10.3390/mi11040376](https://doi.org/10.3390/mi11040376).

RAMESH S., DIAS G., ANDREW J.-J., SAHA S., MAUREL F. & FERRARI S. (2021). Multimodal Web Page Segmentation Using Self-organized Multi-objective Clustering. *ACM Transactions on Information Systems*. HAL : [hal-03332539](https://hal.archives-ouvertes.fr/hal-03332539).

SAFI W., MAUREL F., ROUTOURE J.-M., BEUST P., MOLINA M., SANN C. & GUILBERT J. (2019). Blind Navigation of Web Pages through Vibro-tactile Feedbacks. In *25th ACM Symposium on Virtual Reality Software and Technology (VRST 2019)*, Sydney, Australia. DOI : [10.1145/3359996.3364758](https://doi.org/10.1145/3359996.3364758), HAL : [hal-02309657](https://hal.archives-ouvertes.fr/hal-02309657).

Litte_Bot : le bot qui donne la réplique dans le style Molière

Anna Pappa

Universite Paris 8

apappa@univ-paris8.fr

Le chatbot *Litte_Bot* fait partie d'un projet collaboratif (du même nom : https://eur-artec.fr/projets/litte_bot/) qui transpose un personnage en chatbot, grâce à une intelligence artificielle entraînée sur le corpus des pièces de théâtre du 17^e siècle, incluant la totalité des oeuvres de Molière. Le but de ce projet est de créer un agent conversationnel, capable de donner la réplique comme un personnage d'une pièce de théâtre, avec des traits accentués sur le personnage de Don Juan. Le projet s'inscrit dans la perspective du quatre-centième anniversaire de la naissance de Molière en 2022.

La problématique du projet :

- Construire un dataset uniquement des pièces de théâtre classique à partir des oeuvres de Molière (et quelques autres auteurs du théâtre du 17^e).
- Adapter le modèle d'apprentissage basé sur les dialogues des pièces du théâtre.

Création dataset

Nous avons effectué plusieurs entraînements d'apprentissage en utilisant chaque fois un dataset de plus en plus volumineux. Nous avons commencé avec un dataset de 4k de répliques couples en format *input / target*. Le prototype actuel, en phase de test, est entraîné avec 68k de répliques. Le corpus qui a été utilisé pour créer le dataset est composé principalement de pièces du théâtre de Molière en prose et en vers, ainsi que d'autres pièces des auteurs du théâtre classique français du 17^e comme Dorimond, Rosimond, Villiers et Corneille. Pour avoir un dataset équilibré, afin d'optimiser l'apprentissage par le modèle choisi, il fallait respecter un équilibre entre toutes les répliques couples. Cette tâche a exigé plusieurs étapes de préparation et pré-traitement avant d'avoir la version actuelle du dataset.

- normaliser et nettoyer les textes
- respecter les dialogues entre les personnages
- équilibrer le poids de chaque couple 'input/target'

Exemple : Input : quoi divine isabelle on a donc résolu de vous donner alonce, et vous l'avez voulu ? Target : léandre, mon tuteur me choisit un époux. si j'étais toute à moi, je serais toute à vous.

- automatisation du processus

Modèle Seq2Seq

Pour l'apprentissage, nous avons implementé un modèle génératif sequence à sequence (Seq2Seq). Le choix a été orienté par la capacité de ces modèles à répondre à des tâches non structurées, telles que le dialogue à partir des mots et la génération de réponse à partir de mots, et donc éliminer le besoin de

réponses préparées. Parmi les exemples les plus marqués nous citons les *Microsoft XiaoIce* (Markoff & Mozur, 2015), (Zhou *et al.*, 2018) et *Smart Reply* de Google (Kannan *et al.*, 2016). Dans l'état de l'art des modèles génératifs actuels utilisés dans (Thu Le *et al.*, 2018) on trouve l'architecture d'un réseau neuronal récurrent hiérarchique, où dans un niveau il encode et décode et dans un autre il met à jour un état de conversation. Plus précisément un GRU bidirectionnel encode l'entrée, dont la sortie est utilisée pour mettre à jour un RNN au niveau de la conversation, la sortie de ce RNN est ensuite décodée avec un autre RNN pour produire une réponse. Au prix d'éléments sémantiques plus riches en termes de littérature et du théâtre, le modèle génératif interprète souvent mal les arguments ou produit des réponses donnant l'impression qui manquent de cohérence. Néanmoins, malgré la petite taille du dataset (5 Mo seulement, comparé par exemple au Chatbot Meena de Google qui est entraîné sur le meme modèle mais avec 341 Go de données textuelles), les réponses sont des répliques du théâtre du 17e ayant un style et respectant les rimes.

Conclusion

Nous avons présenté brièvement un prototype d'agent conversationnel qui donne la réplique comme un personnage du théâtre français du 17e siècle. La phase des tests continue et la version corrigée intégrera les éléments suivants : une partie de dialogue extraite par les essais en ligne avec le chatbot, un dataset encore plus volumineux (environ 100k) et plus équilibré (plus homogène au niveau poids pour chaque entrée). Suite à cette phase nous évaluerons le chatbot avec une partie de pièces de théâtre qui servira de tests afin de mesurer la similarité avec le texte original de Molière. Extrait :

Litte_Bot : quel sera cet objet ?

anna : notre conversation pour notre plaisir

Litte_Bot : ces nobles de province y sont un peu fâcheux.

anna : qui donc ? quels nobles ?

Litte_Bot : je l'ai trop peu d'ami ?

anna : je ne sais pas, votre amitié m'importe

Litte_Bot : quelle gloire !

anna : je ne vous suis plus ...

Litte_Bot : vous n'avez pas sujet de lui porter envie.

anna : et vous ? qu'avez-vous à dire ?

Litte_Bot : puisqu'il sait l'aventure, avouez tout, madame.

anna : êtes-vous amoureux ?

Litte_Bot : oui, ma belle. et que voici l'enferp qui s'approche.

Références

KANNAN A., KURACH K., RAVI S., KAUFMANN T., TOMKINS A., MIKLOS B., CORRADO G., LUKACS L., GANEA M., YOUNG P. & RAMAVAJJALA V. (2016). Smart reply : Automated response suggestion for email. DOI : [10.48550/ARXIV.1606.04870](https://doi.org/10.48550/ARXIV.1606.04870).

MARKOFF J. & MOZUR P. (2015). For sympathetic ear, more chinese turn to smartphone program. *NY Times*.

THU LE D., NGUYEN C.-T. & NGUYEN K. A. (2018). Dave the debater : a retrieval-based and generative argumentative dialogue agent. In *Proceedings of the 5th Workshop on Argument Mining*, p. 121–130, Brussels, Belgium : Association for Computational Linguistics. DOI : [10.18653/v1/W18-5215](https://doi.org/10.18653/v1/W18-5215).

ZHOU L., GAO J., LI D. & SHUM H.-Y. (2018). The design and implementation of xiaoice, an empathetic social chatbot. DOI : [10.48550/ARXIV.1812.08989](https://doi.org/10.48550/ARXIV.1812.08989).

Recherche du but de l'utilisateur par une session interactive - jaii2021 : Journée "Accès interactif à l'information" (AFIA-THL / ARIA / GDR TAL)

Agnès Mustar, Sylvain Lamprier, Benjamin Piwowarski

LIP6

agnes.mustar@lip6.fr

RÉSUMÉ

Lorsqu'un utilisateur effectue une recherche d'information complexe, il tâtonne avant de trouver les termes qui le mèneront aux documents pertinents. Pour faciliter cette recherche, plusieurs méthodes existent : la suggestion de requêtes (Sordoni *et al.*, 2015; Dehghani *et al.*, 2017; Mustar *et al.*, 2021), le retour utilisateur implicite (Levine *et al.*, 2017) ou explicite (Yu *et al.*, 2019), les chat-bots (Chen *et al.*, 2019), ...). Ces différentes méthodes présentent des inconvénients : certaines sont coûteuses en temps pour l'utilisateur, d'autres prennent en compte un retour implicite. Nous proposons une méthode qui demande peu de temps à l'utilisateur, tout en permettant d'obtenir un retour explicite, donc plus robuste.

À partir de la question initiale de l'utilisateur, nous lui proposons à chaque étape de la session k questions connexes. L'utilisateur choisit alors celle qui reflète le mieux son besoin d'information. L'historique de préférences de l'utilisateur est utilisé pour trouver son but plus rapidement. On dispose ainsi d'une suite de préférences relatives à un ensemble de questions.

Afin d'évaluer notre modèle sans recours à des utilisateurs réels, nous proposons également une méthode pour obtenir des données interactives en simulant un utilisateur qui effectue des choix.

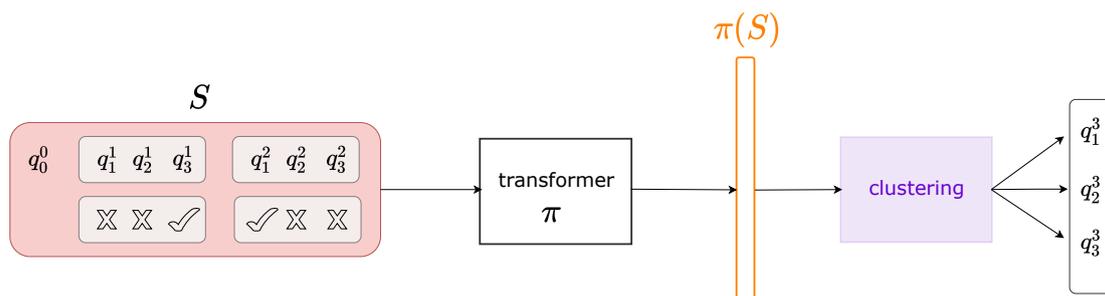


FIGURE 1 – Processus de suggestion de questions

Nous supposons qu'il existe un espace *continu* permettant de représenter l'ensemble des questions possibles. Cet espace est obtenu en encodant l'ensemble des questions avec un transformer pré-entraîné (Hofstätter *et al.*, 2021). La suggestion de questions est alors effectuée suivant ce processus (figure 1) :

- le modèle de suggestions prédit un point de l'espace (des questions) à partir de la question

initiale et de l'historique de préférences ;

- les k questions les plus proches (dans l'espace continu) sont sélectionnées puis sont regroupées (clustering) en N clusters ;
- les centres des N clusters sont utilisés comme questions - l'utilisation du clustering assure la diversité des suggestions.

Le modèle (que l'on notera π) est appris en deux phases (figure 2). La première est une phase d'apprentissage supervisé itératif, où l'on cherche à minimiser la distance du point prédit par le modèle au but final de l'utilisateur (L_π). Pour ce premier apprentissage supervisé, nous générons des séquences de préférences. Ces dernières sont obtenues en échantillonnant d'abord aléatoirement des paires question initiale-but : (q_0^0, g) d'un jeu de données basé sur Wikipedia (Hofstätter *et al.*, 2021). Ensuite, on suit le processus décrit Figure. 1 en utilisant à chaque itération d'apprentissage le modèle π appris à l'itération précédente. Lors de la première itération, un modèle de suggestions aléatoire est utilisé. Au fil des itérations, les trajectoires sont plus riches et le modèle plus performant.

Dans un deuxième temps, une phase d'apprentissage par renforcement (car le processus de clustering n'est pas dérivable) permet d'optimiser directement la distance entre la question choisie par l'utilisateur et son but. On introduit un réseau critique, appris en minimisant L_{critic} (voir figure) qui permet d'estimer la récompense, grâce à laquelle le clustering est pris en compte.

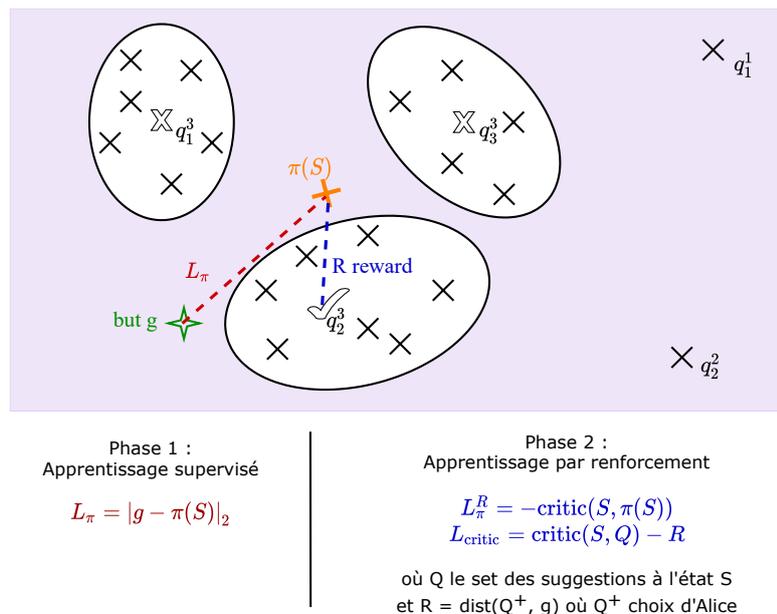


FIGURE 2 – Illustration du processus d'apprentissage du modèle de suggestions

Nous présenterons les résultats préliminaires obtenus par ces modèles en nous intéressant au nombre d'itérations nécessaires pour atteindre le but (s'il est atteint). Les travaux futurs s'intéresseront à rendre plus réaliste notre modélisation, en s'intéressant en particulier à l'utilisateur, ainsi qu'à de nouvelles interactions (par exemple, le choix d'un document).

Références

CHEN C., FU C., HU X., ZHANG X., ZHOU J., LI X. & BAO F. S. (2019). [Reinforcement](#)

[Learning for User Intent Prediction in Customer Service Bots](#). New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/3331184.3331370](#).

DEHGHANI M., ROTHE S., ALFONSECA E. & FLEURY P. (2017). [Learning to Attend, Copy, and Generate for Session-Based Query Suggestion](#). In *Proceedings of the 26th ACM International on Conference on Information and Knowledge Management, CIKM '17*, p. 1747–1756, New York, NY, USA : ACM.

HOFSTÄTTER S., LIN S.-C., YANG J.-H., LIN J. & HANBURY A. (2021). Efficiently Teaching an Effective Dense Retriever with Balanced Topic Aware Sampling. In *Proc. of SIGIR*.

LEVINE N., ROITMAN H. & COHEN D. (2017). [An Extended Relevance Model for Session Search](#). In *The 40th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '17*, p. 865–868, New York, NY, USA : ACM.

MUSTAR A., LAMPRIER S. & PIWOWARSKI B. (2021). [On the study of transformers for query suggestion](#). *ACM Transactions on Information Systems (TOIS)*, **40**(1), 1–27.

SORDONI A., BENGIO Y., VAHABI H., LIOMA C., GRUE SIMONSEN J. & NIE J.-Y. (2015). [A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion](#). In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15*, p. 553–562, New York, NY, USA : ACM.

YU L., CHEN H., WANG S., ARTZI Y. & LEI T. (2019). [Interactive Classification by Asking Informative Questions](#). *arXiv preprint arXiv :1911.03598*.

Renewal - plateforme d'évaluation *online* pour la recommandation d'articles d'actualité

Julien Hay^{a,b,c}, Bich-Liên Doan^{b,c}, Fabrice Popineau^{b,c}
Erik M. Bray^{c,d,e}, Anne-Catherine Letournel^{c,d,e}
Octopeek^a, CentralSupélec^b, LISN^c, CNRS^d, Université Paris-Saclay^e

RÉSUMÉ

Nous proposons une plateforme de recommandation d'articles d'actualité *online*. L'objectif de cette plateforme est double : d'une part, réduire le biais des recommandations *offline* lié à la spécificité des articles d'actualité (volatilité, popularité dynamique, contenu textuel, etc.), d'autre part, fournir un cadre interactif utilisateur-système unique, pour réaliser des compétitions d'algorithmes de recommandation, avec une évaluation par les utilisateurs en temps réel. La plateforme est développée au sein du laboratoire LISN, en phase de bêta-test, et a déjà été utilisée dans un contexte pédagogique.

Mots-clés : Systèmes de recommandation, évaluation *online*, articles d'actualité

1 Renewal : une plateforme interactive d'évaluation *online*

Un système de recommandation d'articles d'actualité vise à réduire la surcharge d'informations à laquelle sont confrontés les utilisateurs en filtrant les articles selon leurs préférences. La recommandation d'articles d'actualité est une tâche spécifique dans le domaine des systèmes de recommandation en raison de la nature des articles (par ex. popularité dynamique, contenu textuel riche, catalogue important et en croissance continue, durée de vie courte) et de la consommation spécifique des lecteurs (par ex. besoin de s'informer sur l'actualité récente, existence d'intérêts à long et à court terme, importance du contexte comme le jour de la semaine et le lieu).

L'évaluation *offline* des systèmes de recommandation est répandue car plus simple à mettre en œuvre. Elle est cependant limitée dans sa capacité à évaluer la réelle satisfaction de l'utilisateur (Bellogín & Said, 2018). CLEF NewsREEL (Hopfgartner *et al.*, 2016) fut historiquement la seule plateforme qui permettait de pallier les biais d'évaluation *offline* en proposant un cadre d'évaluation *online*. L'avantage de cette plateforme était son large panel d'utilisateurs et d'articles qui ont permis de nombreuses publications dans ce domaine.

Cependant, CLEF NewsREEL présentait un certain nombre de limites que nous proposons de surmonter dans la plateforme *Renewal* Hay *et al.* (2020a) :

- mise à disposition d'une application qui permet une recommandation de différentes sources en même temps, rassemblée dans une même page de l'application ;
- prise en compte d'historique long terme par l'enregistrement des utilisateurs (qui peut être anonyme) ;
- introduction de la mesure *click-and-read* qui permet de s'assurer de la fiabilité des retours de pertinence (nous confirmons l'intérêt de lecteur par leur temps de lecture et pas uniquement le clic) ;

- allègement de contraintes dans l’implémentation des systèmes (par ex. temps de réponse, disponibilité des données, mise en cache de recommandations) et facilitation par la mise à disposition d’exemples et d’une API simple d’utilisation.

2 Expérimentation

Notre première compétition *Renewal* s’est déroulée de juin à décembre 2021 et impliquait trois systèmes implémentés par trois équipes d’étudiants. Nous avons également mis en jeu trois baselines (une première reposant sur les mots-clés des articles, une seconde effectuant une recommandation aléatoire et une dernière reposant sur la popularité des articles). Nous avons demandé à 20 utilisateurs de lire des articles sur l’application mobile *Renewal* durant quelques semaines afin de permettre l’évaluation de ces systèmes. À chaque fois qu’un utilisateur actualisait sa liste de recommandation, nous entrelaçons environ 30 articles recommandés par deux systèmes, et l’utilisateur pouvait ainsi interagir avec les articles présentés (*click-and-read*). Les interactions utilisateur-système nous permettent de collecter des retours de pertinence pour chaque lecteur. Nous collectons les *click-and-reads* et déterminons quel système est vainqueur par la méthode *Team Draft* (Radlinski *et al.*, 2008). Nous calculons ensuite un score Elo sur la base de toutes les confrontations observées. Enfin, nous affichons un classement des systèmes. Une description plus complète de notre plateforme ainsi qu’une simulation de compétition montrant sa faisabilité a été décrite dans l’article Hay *et al.* (2020a) et la thèse Hay (2021).

3 Conclusion

Dans ce travail, nous avons montré l’intérêt et la faisabilité de *Renewal* en organisant une compétition à petite échelle. À terme, nous souhaitons proposer le co-développement de nouvelles fonctionnalités de la plateforme *Renewal* à la communauté TALN et RI, et prévoyons l’organisation de compétitions d’algorithmes de recommandation d’articles d’actualité (en confrontant notamment le modèle de préférences stylistiques développé par Hay *et al.* (2020b) à d’autres modèles) à grande échelle, dans le cadre d’un workshop.

Références

- BELLOGÍN A. & SAID A. (2018). Recommender systems evaluation. In *Encyclopedia of Social Network Analysis and Mining* :. 2 édition. DOI : [10.1007/978-1-4939-7131-2_110162](https://doi.org/10.1007/978-1-4939-7131-2_110162).
- HAY J. (2021). *Apprentissage de la représentation du style écrit, application à la recommandation d’articles d’actualité*. Thèse de doctorat.
- HAY J., DOAN B.-L., POPINEAU F. & AIT ELHARA O. (2020a). *Renewal* : an online competition platform for news recommender systems. In *Proceedings of Challenges in Machine Learning at NeurIPS 2020*.
- HAY J., DOAN B.-L., POPINEAU F. & AIT ELHARA O. (2020b). Representation learning of writing style. In *Proceedings of the 6th Workshop on Noisy User-generated Text (W-NUT 2020)*.
- HOPFGARTNER F., BRODT T., SEILER J., KILLE B., LOMMATZSCH A., LARSON M., TURRIN R. & SERÉNY A. (2016). Benchmarking news recommendations : The clef newsreel use case. *SIGIR Forum*, **49**(2), 129–136. DOI : [10.1145/2888422.2888443](https://doi.org/10.1145/2888422.2888443).
- RADLINSKI F., KURUP M. & JOACHIMS T. (2008). How does clickthrough data reflect retrieval quality ? In *CIKM ’08*.

SigLa - Une plateforme de développement d'animations en langue des signes

Bastien David, Jonathan Mutal, Irene Strasly, Johanna Gerlach & Pierrette Bouillon

Faculté de Traduction et d'Interprétation, Université de Genève

Bastien.David@unige.ch, Jonathan.Mutal@unige.ch, Irene.Strasly@unige.ch,
Johanna.Gerlach@unige.ch & Pierrette.Bouillon@unige.ch

RÉSUMÉ

Mots-clés : Communication accessible - Langue des signes - Avatar - Médecine

À notre époque, une information accessible doit être impérativement proposée aux minorités qui en font la demande, notamment aux personnes avec un handicap auditif et dont la langue des signes (LS) est la langue maternelle. 80% des personnes sourdes ont en effet des difficultés de lecture et l'utilisation de la LS est indispensable pour assurer une bonne communication ([World Federation of the Deaf, 2003](#)).

Dans ce contexte, le système [BabelDr \(Bouillon et al., 2021\)](#) s'intéresse à l'accessibilité de l'information médicale et vise à traduire automatiquement les questions et les instructions médicales pour les minorités linguistiques vivant dans le canton de Genève, dont les personnes sourdes ([Strasly et al., 2018](#)). Pour assurer la fiabilité des traductions, le système ne peut traduire que les phrases de sa mémoire de traduction (environ 10'000 phrases canoniques). Ce dernier inclut cependant un modèle de reconnaissance vocale qui permet au médecin de s'exprimer librement à l'oral. Il lie ensuite par des techniques neuronales le résultat de la reconnaissance vocale à la phrase pré-traduite la plus proche de la mémoire. Si celle-ci est validée par le médecin, elle sera oralisée / signée pour le patient. Les réponses du patient sont non verbales ou avec des pictogrammes ([Norré et al., 2021](#)).

Étant donné que la LS est une langue visuelle, non-écrite et tri-dimensionnelle, la traduction de ce type de corpus en LS présente différentes difficultés, par rapport aux langues écrites. Une solution est l'enregistrement de vidéos par des interprètes humains ([Strasly et al., 2018](#)). Ce procédé, même s'il produit des traductions de très bonne qualité, est coûteux et pose rapidement des problèmes de gestion des vidéos et de portabilité à de nouvelles données. L'emploi de personnages virtuels propose une alternative intéressante pour produire des animations en LS de manière plus productive.

Dans cette présentation, nous décrivons la plateforme en ligne SigLa qui permet de générer des animations avec des personnages virtuels JASigning, complètement synthétisés ([Ebling & Glauert, 2013](#)). Les animations sont produites à partir de la notation Gestural SiGML ([Elliott et al., 2004](#)). Ce code est lui-même généré à partir de différentes ressources linguistiques, dont un glossaire qui décrit les signes manuels avec le langage HamNoSys ([Prillwitz et al., 1989](#)) et une grammaire qui modélise le lien entre un ensemble de phrases sources et leur représentation en langue des signes (table de signes), avec les différents éléments manuels et non manuels nécessaires pour produire le code G-SiGML ([Rayner et al., 2015](#)).

La plateforme offre plusieurs fonctionnalités. Lors du développement des animations, elle permet de stocker et centraliser les ressources linguistiques selon leur nature : le glossaire sous l'onglet `Glossary` (voir Fig. 1) et la grammaire sous `Grammar`. Celles-ci peuvent être ensuite directement

RAYNER M., ARMANDO A., BOUILLON P., EBLING S., GERLACH J., HALIMI S., STRASLY I. & TSOURAKIS N. (2015). Helping Domain Experts Build Phrasal Speech Translation Systems. In J. F. QUESADA, F.-J. MARTÍN MATEOS & T. LOPEZ-SOTO, Édts., *Future and Emergent Trent in Language Technology : First International Workshop (FETLT15)*, volume 9577, p. 41–52. Sevilla (Spain) : Springer. DOI : [10.1007/978-3-319-33500-1_4](https://doi.org/10.1007/978-3-319-33500-1_4).

STRASLY I., SEBAÏ T., RIGOT E., MARTI V., GONZALEZ J. M., GERLACH J., SPECHBACH H. & BOUILLON P. (2018). Le projet BabelDr : rendre les informations médicales accessibles en Langue des Signes de Suisse Romande (LSF-SR). In *Proceedings of the 2nd Swiss Conference on Barrier-free Communication : Accessibility in educational settings (BFC 2018)*, p. 92–96, Geneva (Switzerland).

WORLD FEDERATION OF THE DEAF (2003). Position Paper regarding the United Nations Convention on the Rights of People with Disabilities. <https://www.un.org/esa/socdev/enable/rights/contrib-wfd.htm>.

Simplification automatique de textes scientifiques : SimpleText Lab à CLEF-2022

Liana Ermakova

Université de Bretagne Occidentale, HCTI - EA 4249, France

liana.ermakova@univ-brest.fr

Sílvia Araújo

University of Minho, Portugal

Patrice Bellot

Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

Radia Hannachi

Université de Bretagne Sud, HCTI - EA 4249, France

Stéphane Huet

Avignon Université, LIA, France

Jaap Kamps

University of Amsterdam, Amsterdam, The Netherlands

Elise Mathurin

Université de Bretagne Occidentale, HCTI - EA 4249, France

Diana Nurbakova

Institut National des Sciences Appliquées de Lyon, LIRIS UMR 5205 CNRS, Lyon, France

Irina Ovchinnikova

Sechenov University, Moscow, Russia

Eric SanJuan

Avignon Université, LIA, France

RÉSUMÉ

Le web et les médias sociaux ont démocratisé le partage de l'information et sont devenus la principale source d'information des citoyens, au risque que les utilisateurs se fient à des informations superficielles dans des sources privilégiant les incitations commerciales ou politiques plutôt que l'exactitude et la valeur informative. Les non-experts ont tendance à éviter les sources telles que la littérature scientifique en raison de son langage complexe, de son vernaculaire interne ou de leurs manques d'acculturation scientifique. Les approches de simplification des textes promettent de lever certains de ces obstacles. La campagne d'évaluation SimpleText à la conférence CLEF¹ (Conference and Labs of the Evaluation Forum) 2022 aborde les défis des approches de simplification du texte dans le contexte de la promotion de l'accès à l'information scientifique, en fournissant des données et des métriques appropriés, et en créant une communauté de chercheurs en TAL et en RI travaillant ensemble pour résoudre l'un des plus grands défis d'aujourd'hui. La campagne d'évaluation SimpleText utilisera un corpus de résumés de littérature scientifique et d'articles de vulgarisation scientifique. Elle comporte trois tâches :

- **Tâche 1 (sélection du contenu) : What is in, or out ?** À partir d'une requête, sélectionnez les passages à inclure dans un résumé simplifié.

1. <https://clef2022.clef-initiative.eu/index.php>

- **Tâche 2 (complexity spotting) : What is unclear ?** À partir d'un passage et d'une question, trouver les termes/concepts qui doivent être expliqués pour comprendre ce passage (définitions, contexte, applications,...).
- **Tâche 3 (simplification du texte) : Rewrite this !** A partir d'une requête, simplifiez des passages de résumés scientifiques.
- **Tâche 4 : non partagée.** Toutes soumissions qui utilise nos données sont bienvenues !

Vous pouvez trouver tous les détails concernant notre campagne d'évaluation sur le site de SimpleText : <http://simpletext-project.com/>. Aidez-nous à rendre les résultats scientifiques vraiment accessibles et compréhensibles !

Avoir des connaissances scientifiques est une aptitude importante pour les gens. C'est l'une des clés de l'esprit critique, de la prise de décision objective et du jugement de la validité et de la signification des résultats et des arguments, qui permet de discerner les faits de la fiction. Ainsi, posséder des connaissances scientifiques de base peut également contribuer à préserver sa santé, tant physiologique que mentale. La pandémie de COVID-19 en est un bon exemple. Comprendre le problème lui-même, connaître et appliquer les règles de distanciation sociale et les politiques sanitaires, choisir d'utiliser ou d'éviter tel ou tel traitement ou procédure de prévention peut devenir crucial. Dans le contexte d'une pandémie, l'information qualifiée et opportune doit atteindre tout le monde et être accessible. C'est ce qui motive des projets tels qu'[EasyCovid](#).

Cependant, les textes scientifiques sont souvent difficiles à comprendre car ils nécessitent de solides connaissances de base et utilisent une terminologie délicate. Bien qu'il y ait eu quelques efforts récents sur la simplification des textes (par exemple ([Maddela et al., 2021](#))), l'élimination automatique de ces barrières de compréhension entre les textes scientifiques et le grand public reste un défi ouvert. **SimpleText Lab** rassemble des chercheurs et des praticiens travaillant sur la génération de résumés simplifiés de textes scientifiques. Il s'agit d'un nouveau laboratoire d'évaluation qui fait suite à l'atelier SimpleText-2021 ([Ermakova et al., 2021](#)). Tous les points de vue sur la vulgarisation scientifique automatique sont les bienvenus, y compris, mais sans s'y limiter, les domaines suivants : le traitement automatique des langues (TAL), la recherche d'information (RI), la linguistique, le journalisme scientifique, etc.

SimpleText fournit des données et des repères pour la discussion des défis de la simplification automatique de texte. Nous présenterons l'état de l'art du problème de la simplification automatique des textes scientifiques et décrivons les tâches envisagées par la campagne d'évaluation SimpleText-2022 à CLEF.

Remerciements. L'atelier SimpleText est labélisé par le GDR MADICS (Masses de Données, Informations et Connaissances en Sciences)².

Références

ERMAKOVA L., BELLOT P., BRASLAVSKI P., KAMPS J., MOTHE J., NURBAKOVA D., OVCHINIKOVA I. & SANJUAN E. (2021). Text Simplification for Scientific Information Access : CLEF 2021 SimpleText Workshop. In *Advances in Information Retrieval - 43rd European Conference on {IR} Research, {ECIR} 2021, Lucca, Italy, March 28 – April 1, 2021, Proceedings*, Lucca, Italy.

MADDELA M., ALVA-MANCHEGO F. & XU W. (2021). Controllable Text Simplification with Explicit Paraphrasing. *arXiv :2010.11004 [cs]*. arXiv : 2010.11004.

2. <https://www.madics.fr/ateliers/simpletext/>

Un agent conversationnel pour améliorer la découvrabilité des données ouvertes - jaii2021 : Journée "Accès interactif à l'information" (AFIA-THL / ARIA / GDR TAL)

Anthony Gigerich Mathieu Morey

Dataactivist, 7 bis avenue Saint-Jérôme, 13100 Aix-en-Provence, France

{anthony, mathieu}@dataactivist.coop

Le mouvement de l'ouverture de leurs données par des acteurs publics et privés se traduit par la publication sur des portails web d'un grand nombre jeux de données de taille, de nature et de domaines très divers. L'accès aux jeux de données et leur utilisation sont cependant entravés par différents obstacles socio-techniques (Beno *et al.*, 2017). En premier lieu, la capacité d'un utilisateur à trouver les jeux de données qu'il cherche est déterminée en partie par la capacité d'un jeu de données à être trouvé ou découvert, c'est-à-dire sa découvrabilité (Dasgupta *et al.*, 2007).

Les méthodes d'indexation et de recherche développées pour les contenus web (texte, image, vidéo) et mises en œuvre dans les moteurs de recherche n'étant pas adaptées aux jeux de données ouvertes, structurées (relationnelles, géospatiales) ou non structurées (Neumaier *et al.*, 2016), les fonctions de recherche proposées par les portails de données ouvertes se limitent essentiellement à un calcul de similarité textuelle entre la requête de l'utilisateur et les métadonnées renseignées par les producteurs lors du dépôt : titre, description, mots-clés, auteur... Or, les producteurs décrivent leurs données dans leur contexte de collecte et d'exploitation, avec leur terminologie, alors que les utilisateurs de portails formulent leur besoin de données dans le contexte de leurs propres cas d'usages et avec leur propre terminologie. La recherche sur les portails de données ouvertes souffre donc d'un très fort phénomène de *vocabulary mismatch*, d'autant plus difficile à compenser que les métadonnées (y compris la description) sont souvent très succinctes.

Dans ce contexte, nous avons développé un agent conversationnel pour les portails de données ouvertes, qui aide l'utilisateur à formuler sa requête et recueille ses retours sur la pertinence des jeux de données qui lui sont proposés. Les mécanismes génériques de l'agent conversationnel sont fournis par le framework open source Rasa¹. L'expansion de requête repose sur des ressources lexicales, actuellement des plongements de mots word2vec (Mikolov *et al.*, 2013) appris sur le corpus frWaC (Baroni *et al.*, 2009; Fauconnier, 2015), qui permettaient une amélioration du rappel supérieure à fastText (Bojanowski *et al.*, 2016) et WordNet (WOLF) (Sagot & Fišer, 2008) dans nos essais préliminaires. L'intégration des ressources distributionnelles word2vec et fastText dans un système interactif (nécessitant une latence faible) est permis par la bibliothèque Magnitude (Patel *et al.*, 2018b,a).

Nous travaillons à améliorer le reclassement des résultats, l'exploitation des retours de pertinence et l'intégration de terminologies en complément des ressources lexicales génériques. Notre agent conversationnel est actuellement déployé sur le portail de données ouvertes *Datasud*², dans le cadre d'une expérimentation avec la Région Sud.

1. <https://github.com/RasaHQ/rasa/>

2. <https://www.datasud.fr/>

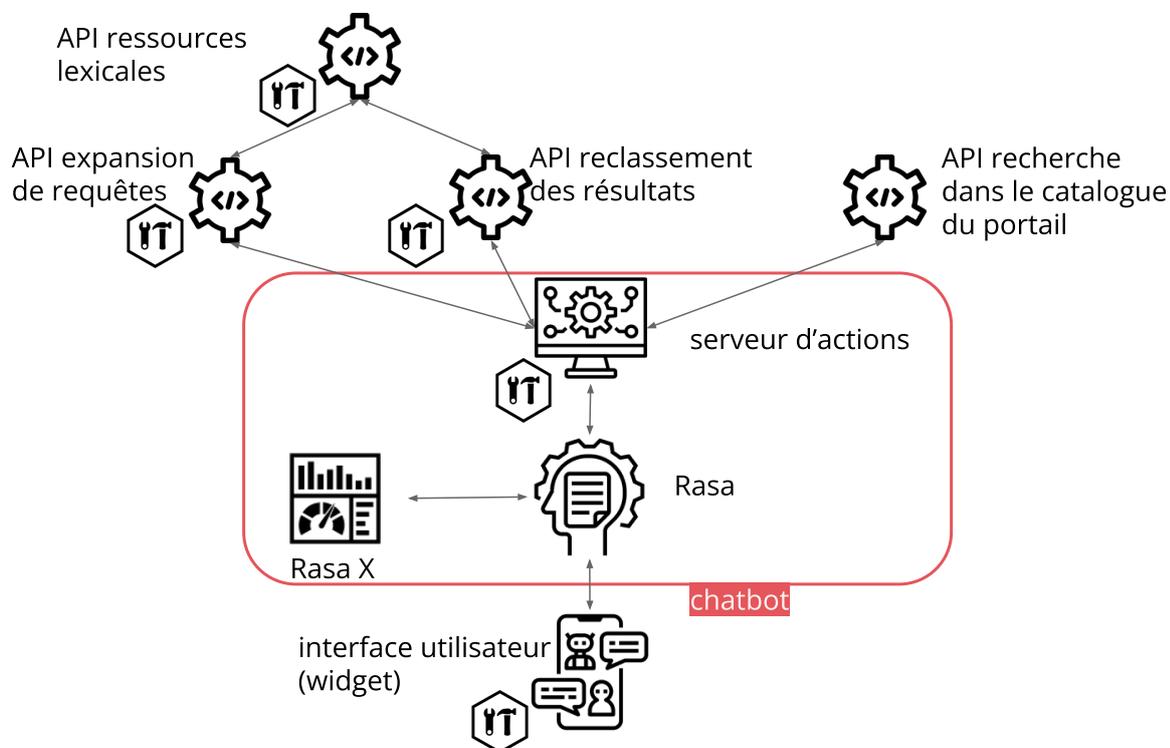


FIGURE 1 – Architecture de l’agent conversationnel

La figure 1 représente l’architecture de notre agent conversationnel. L’usager voit apparaître sur la page d’accueil du portail une interface conversationnelle sous forme d’un *widget*, qui affiche les messages émis par le composant Rasa et transmet à ce dernier les messages saisis par l’usager. Certains messages déclenchent des actions, définies dans le serveur d’actions de Rasa, qui appellent à leur tour des APIs pour l’expansion de requêtes, la recherche de jeux de données dans le catalogue du portail et le reclassement des résultats, avant de les présenter à l’usager via le widget. L’une des actions permet de recueillir le retour de l’usager sur les termes pertinents, une autre action son retour sur les jeux de données pertinents pour sa recherche. Les APIs d’expansion de requêtes et de reclassement des résultats appellent l’API de ressources lexicales pour obtenir des termes similaires et des scores de similarité. L’interface Rasa X permet de consulter l’historique des conversations, y compris les décisions prises par les classifieurs de détection d’intentions et d’entités, d’annoter les conversations avec des étiquettes, de ré-entraîner des modèles pour les classifieurs et de changer de modèles.

Références

- BARONI M., BERNARDINI S., FERRARESI A. & ZANCHETTA E. (2009). The wacky wide web : a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, **43**(3), 209–226.
- BENO M., FIGL K., UMBRICH J. & POLLERES A. (2017). Perception of key barriers in using and publishing open data. *JeDEM-eJournal of eDemocracy and Open Government*, **9**(2), 134–165.
- BOJANOWSKI P., GRAVE E., JOULIN A. & MIKOLOV T. (2016). Enriching word vectors with subword information. *arXiv preprint arXiv :1607.04606*.

- DASGUPTA A., GHOSH A., KUMAR R., OLSTON C., PANDEY S. & TOMKINS A. (2007). The discoverability of the web. In *Proceedings of the 16th international conference on World Wide Web*, p. 421–430 : ACM.
- FAUCONNIER J.-P. (2015). French word embeddings.
- MIKOLOV T., SUTSKEVER I., CHEN K., CORRADO G. S. & DEAN J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, p. 3111–3119.
- NEUMAIER S., UMBRICH J. & POLLERES A. (2016). Automated quality assessment of metadata across open data portals. *Journal of Data and Information Quality (JDIQ)*, **8**(1), 2.
- PATEL A., ALEX, ADMIN P. & JONNY (2018a). plasticityai/magnitude : Release 0.1.120. DOI : [10.5281/zenodo.1607574](https://doi.org/10.5281/zenodo.1607574).
- PATEL A., SANDS A., CALLISON-BURCH C. & APIDIANAKI M. (2018b). Magnitude : A fast, efficient universal vector embedding utility package. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing : System Demonstrations*, p. 120–126.
- SAGOT B. & FIŠER D. (2008). Building a free french wordnet from multilingual resources. In *OntoLex*.

