

Laure Soulier - Maître de conférences

Sorbonne Université

Laboratoire d'Informatique de Paris 6 (LIP6 - UMR 7606 CNRS)

Equipe Machine Learning and Information Access (MLIA)

4 place Jussieu - 75252 Paris Cedex 05

☎ : (+33) 1 44 27 74 91

✉ : laure.soulier@lip6.fr

🌐 : <https://mlia.lip6.fr/soulier/>

🐦 : @LaureSoulier



Fonctions exercées

Depuis Septembre 2015

Maître de Conférences en Informatique

Sorbonne Université, LIP6 UMR 7606 CNRS, UFR 919 Ingénierie

Octobre 2014 – Août 2015

Attachée Temporaire d'Enseignement et de Recherche (1/2 ATER)

Université Toulouse III, IRIT, Faculté des Sciences et d'Ingénierie

Octobre 2011 – Septembre 2014

Doctorante Contractuelle Chargée d'Enseignement (DCCE)

Université Toulouse III, Faculté des Sciences et d'Ingénierie

Diplôme universitaire de 3^e cycle

Octobre 2011 – Décembre 2014

Doctorat en informatique - Boursière du MESR

Définition et évaluation de modèles de recherche d'information collaborative basés sur les compétences de domaine et les rôles des utilisateurs (soutenue le 01/12/2014)

Université Toulouse III - IRIT - Equipe Systèmes d'Informations Généralisés

- *Directrice de thèse* : Pr. Lynda Tamine-Lechani
- *Co-encadrante* : MCF Wahiba Bahsoun
- *Rapporteurs* : Pr. Eric Gaussier (Université Grenoble 1), Pr. Jian-Yun Nie (Université Montréal, Canada)
- *Examineurs* : Pr. Florence Sedes (UPS), Pr. Maarten de Rijke (Université d'Amsterdam), Pr. Patrick Gallinari (Sorbonne Université), Pr. Chihab Hanachi (Université Toulouse I)

Rapport d'activité de recherche et résultats majeurs

Thématiques de recherche et contributions majeures

Ma recherche est motivée par la proposition de nouveaux modèles pour la recherche d'information (2011-2018) et le traitement automatique du langage naturel (2015-2018). L'objectif commun de ces modèles est de faciliter le traitement de données textuelles, éventuellement combinées à d'autres sources d'informations, telles que les logs utilisateurs (2011-2018), les données sociales (2014-2015), les ressources sémantiques (2015-2018) ou encore les données visuelles (2015-2018). J'ai également contribué à l'application de modèles formels dans des domaines particuliers (la recherche de produits [Jab+16b; Jab+15; Jab+16a] ou la cuisine [Car+18b; Car+18a]). Dans un premier temps, ma recherche s'est intéressée à la formalisation de modèles basés sur des techniques de RI (pondération des termes, modèles d'appariement, ...). Aujourd'hui, mes travaux sont focalisés sur l'analyse/l'exploitation de données textuelles à partir de méthodes innovantes reposant sur l'apprentissage profond (*Deep learning*). Ci-dessous, mes thématiques de recherche :

- **Recherche d'information collaborative (2011-2016) et socio-collaborative (2014-2015)**
- **Recherche d'information sémantique basée sur l'apprentissage profond (2015-2018)**
- **Apprentissage de représentation multimodale (2015-2018)**
- **Data-to-text : génération de synthèses à partir de données structurées (initié en 2018)**
- **Systèmes de dialogue pour la recherche d'information (initié en 2018)**

Je présente dans ce qui suit mes travaux effectués ces cinq dernières années.

Recherche d'information sémantique basée sur l'apprentissage profond (2015-2018)

Je m'intéresse depuis 2015 au développement de méthodes d'apprentissage profond pour la recherche d'information. Plus particulièrement, j'ai focalisé mon attention sur une thématique innovante en 2015 ayant pour objectif de combiner la sémantique distributionnelle capturée par les réseaux de neurones (e.g., plongements de mots ou *word embeddings*) avec la sémantique relationnelle recensée dans les sources de connaissances externes (e.g., les ontologies ou ressources sémantiques). Ce travail a été initié lors d'une thèse (G.-H. Nguyen) que j'ai co-encadré en collaboration avec l'IRIT à Toulouse. Nous avons proposé deux principaux modèles : une approche de représentation hybride (sémantique distributionnelle et relationnelle) des documents afin d'injecter ces représentations dans des modèles de recherche d'information [Ngu+17c; Ngu+18a; Ngu+18b] ainsi qu'un modèle neuronal de recherche d'information permettant d'apprendre simultanément la représentation hybride des documents ainsi que la fonction d'ordonnement des documents [Ngu+17a; Ngu+17b].

Apprentissage de représentation multimodale (2015-2018)

J'ai également enrichi mes travaux de recherche autour de l'apprentissage de représentation en considérant une autre source d'information complémentaire au texte : les documents visuels ou vidéos. Cette thématique émerge depuis 2014 avec les récentes avancées de l'apprentissage profond en vision, en plus récemment autour du texte. Dans le cadre du projet Chistera MUSTER et également d'un co-encadrement de thèse (E. Zablocki), nous avons proposé un axe innovant pour l'apprentissage de représentation de mots en considérant dans la modalité visuelle son apparence mais également son contexte (à savoir les objets qui l'entourent) [Zab+18]. Nous avons poursuivi ces travaux selon deux axes : la détection d'objets dans une image à partir d'un texte [Zab+19] ainsi que l'apprentissage de représentation de phrases à partir du contexte multimodal [Bor+19a]. J'ai également collaboré avec Mathieu Cord, spécialisé vision de notre équipe pour développer un modèle d'apprentissage multimodal de recettes de cuisine permettant à la fois de rechercher des recettes dans un cadre de recherche d'information ou également de classifier les recettes [Car+18b; Car+18a].

Faits marquants : Participation au projet Chistera MUSTER et co-organisation du workshop DECOR à ICDE en 2018.

Data-to-text : génération de synthèses à partir de données structurées (initié en 2018) Depuis 2018, je me suis intéressée à la génération de textes, plus particulièrement contrainte par des données structurées (tableaux, graphes de connaissances, triplets RDF, séries numériques...). Cette thématique est appelée Data-to-Text et connaît un engouement ces dernières années, tant au niveau scientifique grâce au développement des modèles de langue, qu'au niveau industriel pour ses nombreuses applications dans les domaines de la finance, de l'assurance, du médical, ... Nos contributions dans ce domaine sont regroupées en plusieurs axes : 1) la prise en compte de la structure des données avec

la proposition d'un modèle hiérarchique [Reb+20a; Reb+20b], 2) la gestion des hallucinations dans les textes générés avec la proposition de modèles [Reb+20c; Reb+21a] et de métriques d'évaluation [Reb+21b]. Ces travaux réalisés dans le cadre d'une thèse ont mis en évidence de nombreux enjeux que nous souhaitons aborder dans les prochaines années : apprentissage sur des données non-alignées, l'intégration de modèles de raisonnement numériques (e.g., comparaison, réalisation de calculs tels que la moyenne), transfert à des domaines particuliers (biologie et "document intelligence"). *Faits marquants* : Participation à l'ANR PRCE ACDC qui débutera en avril 2022 sur la problématique de raisonnement numérique en collaboration avec le MNHN, Recital et le LAMSADE.

Systèmes de dialogue pour la recherche d'information (initié en 2018) Je me suis également investie dans une nouvelle thématique de recherche initiant un nouveau paradigme en recherche d'information basé sur les systèmes conversationnels. Pour réaliser ces travaux, j'ai obtenu une ANR JCJC SESAMS qui s'intéresse à la compréhension du besoin en information dans des systèmes conversationnels et à la proposition de modèles proactifs anticipant les besoins de l'utilisateur afin de l'aider dans sa recherche. Nos contributions passés ou en cours sur ce projet portent sur des sous-tâches particulières d'un système conversationnel pour la recherche d'information (à savoir, la reformulation de requêtes [Ais+18], avec également la participation à TREC Cast 2021¹, la suggestion interactive de requêtes [Erb+21] et la génération de réponses à partir d'un besoin en information). En complément à ce projet, je m'intéresse depuis deux ans à la thématique d'apprentissage continu ("lifelong learning") qui est un point essentiel pour les systèmes conversationnels pour la recherche d'information, où l'adaptation aux multiples utilisateurs et aux requêtes posées est primordiales. Les enjeux d'application aux systèmes conversationnels étant nombreux, nous avons dans un premier temps choisi de nous focaliser sur des tâches de recherche d'information, plus simples. Cette thématique a été initiée dans le cadre d'une collaboration avec l'IRIT (équipe IRIS) lors d'un stage d'été 2020 qui a donné lieu à une publication à ECIR 2021 [Lov+21b]. Un postdoctorant a travaillé également dans la continuité de ce stage [Ger+22]. Nous avons également travaillé sur les problématiques de génération de réponse pour la RI conversationnelle [Dje+22]. Un doctorant a également débuté en avril 2021 sur la formalisation de modèles d'apprentissage par renforcement dans un contexte d'apprentissage continu. Cette thèse est principalement théorique et n'aborde pas les systèmes conversationnels [Gay+22]. Trois travaux sont actuellement en soumission pour cette thématique.

Faits marquants : Le projet ANR JCJC SESAMS dont je suis porteuse, en collaboration avec FAIR, CRISAL/Inria Lille, et le DIRO/Université de Montréal.

Encadrement de thèses

Maria Lebedeva - décembre 2021

Ecole Doctorale : EDITE

Titre : Data-to-text and text-to-data generation models

Encadrement à 33% Laure Soulier - 33% Vincent Guigue - 33% Patrick Gallinari

Tristan Luiggi - septembre 2021

Ecole Doctorale : EDITE - Thèse CIFRE Upskills

Titre : Génération de textes pour des données structurées : raisonnement numérique et structuration des données

Encadrement à 50% Laure Soulier - 50% Vincent Guigue

Jean-Baptiste Gaya - avril 2021

Ecole Doctorale : EDITE - Thèse CIFRE FAIR

Titre : Apprentissage continu pour l'apprentissage par renforcement

Encadrement à 50% Laure Soulier - 50% Ludovic Denoyer

Liste des publications : [Gay+22]

Pierre Erbacher - janvier 2021 dans le cadre de l'ANR JCJC

Ecole Doctorale : EDITE

1. <http://www.treccast.ai/>

Titre : Recherche d'information conversationnelle
Encadrement à 50% Laure Soulier - 50% Ludovic Denoyer
Liste des publications : [Erb+21]

Clément Rebuffel - octobre 2018 - soutenance prévue en mai 2022

Ecole Doctorale : EDITE - Thèse CIFRE BNP Paribas
Titre : Modèle de génération de résumés à partir de données structurées (Data-to-text)
Encadrement à 70% Laure Soulier - 30% Patrick Gallinari
Liste des publications : [Reb+21a; Reb+21b; Reb+20c; Reb+20a; Reb+20b; Reb19], soumission en cours à TACL

Eloi Zablocki - octobre 2016 - soutenance en octobre 2019

Ecole Doctorale : EDITE
Titre : Apprentissage de représentation multimodale basée sur le contexte visuel
Encadrement à 33% Laure Soulier - 33% Benjamin Piwowarski - 33% Patrick Gallinari
Liste des publications : [Zab+17; Zab+18; Zab+19; Bor+19a; Bor+19b]

Gia-Hung Nguyen - octobre 2015 - soutenance en décembre 2018

Ecole Doctorale : MITT (Toulouse)
Titre : Modèle neuronal pour la recherche d'information sémantique
Encadrement à 40% Laure Soulier - 40% Lynda Tamine (IRIT) - 20% Nathalie Souf (IRIT)
Liste des publications : [Ngu+16; Ngu+17a; Ngu+17b; Ngu+17c; Ngu+18a; Ngu+18b], article en cours de rédaction pour TKDE

Encadrement de post-doctorants

2021 Thomas Gérald (1 an)
 — Titre : Apprentissage continu pour la recherche d'information
 — Période : janvier à décembre 2021
 — Résultat : [Ger+22], [Dje+22], TREC CAST participation

Encadrement de stagiaires

2021 Hanane Djeddal (co-encadrement avec l'IRIT)
 — Titre : Génération de réponses pour la recherche d'information complexe
 — Période : février à août 2021
 — Résultat : [Dje+22]
 Nada Staouite (collaboration avec Critéo)
 — Titre : Data-to-text / Text-to-Data avec des données non alignées
 — Période : juin à novembre 2021
 Nawel Astouati
 — Titre : Participation TREC CAST
 — Période : juin à août 2021
 — Résultat : Participation TREC CAST

- 2020 Jesus Louvon (stagiaire toulousain - co-encadrement avec l'IRIT)
— Titre : Lifelong learning et recherche d'information neuronale
— Période : avril à septembre 2020
— Résultat : article ECIR 2021 [Lov+21b]
Amine Djeghri et Aymen Merrouche (stagiaire M1 DAC)
— Titre : Modèle de langue contrôlable - paramétrage et évaluation / Systèmes conversationnels pour la recherche d'information
— Période : juin à août 2020
- 2019 Adrien Pouyet (stagiaire M2 DAC - co-encadrement avec Ludovic Denoyer)
— Titre : Système conversationnel pour la recherche d'information
— Période : mars à août 2019
Ismaël Bonneau et Yannis Karmim (stagiaire M1 DAC)
— Titre : Apprentissage de représentation textuelle basé sur les ressources sémantiques
— Période : juin à août 2019
- 2018 Wafa Aissa (stagiaire M2 DAC - co-encadrement avec Ludovic Denoyer)
— Titre : Système conversationnel pour la recherche d'information
— Période : mars à août 2018
— Résultat : article accepté à SCAI@EMNLP 2018 [Ais+18]
Clément Bosc (stagiaire M1 DC à Toulouse - co-encadrement avec Lynda Tamine/IRIT)
— Titre : Apprentissage de représentation textuelle basé sur les ressources sémantiques
— Période : mai à juillet 2018
— Résultat : article de journal Document Numérique [Ngu+20]
- 2017 Patrick Bordes (stagiaire M2 MVA - co-encadrement avec Benjamin Piwowski et Patrick Gallinari)
— Titre : Extraction de relations à partir de données multimodales
— Période : février à juillet 2017
— Résultat : participation à une campagne d'évaluation CLEF 2017 [Zab+17]
Thomas Gandillon (stagiaire M1 DC à Toulouse - co-encadrement avec Lynda Tamine/IRIT)
— Titre : Modèle neuronal d'interaction pour la RI sémantique
— Période : avril à juillet 2017
- 2016 Thomas Gérald (stagiaire M2 DAC - co-encadrement avec Benjamin Piwowski et Patrick Gallinari)
— Titre : Classification de relation temporelles
— Période : février à juillet 2016
- 2015 Gia-Hung Nguyen (stagiaire M2 RIBD à Toulouse- co-encadrement avec Lynda Tamine)
— Titre : Recherche d'information collaborative sur les réseaux sociaux
— Période : février à août 2015
— Résultat : article CIKM 2016 [Lau+16b]
- 2014 Laurent Clastrier (stagiaire M1 Informatique à Toulouse- encadrement seule)
— Titre : Analyse de logs de recherche d'information collaborative
— Période : avril à juillet 2014
— Résultat : article CIKM 2015 [Tam+15]
Paul Mousset (stagiaire L3 - co-encadrement avec Lamjed Ben Jabeur et Lynda Tamine)
— Titre : Analyse exploratoire de trace de collaboration sur les réseaux sociaux
— Période : avril à juillet 2014
— Résultat : article HT 2016 [Tam+16c]

Publications parues ou acceptées au cours des quatre années précédentes

Le tableau suivant synthétise mes contributions scientifiques publiées dans des conférences/journaux internationaux et nationaux. Je précise le rangs Core² des conférences et des journaux internationaux ainsi que l'Impact Factor disponible sur le site de l'éditeur. Un tableau des synthèses des publications est fourni à la page suivante.

Type de publication	Nombre
Articles de journaux internationales	6
Articles de journaux nationaux	2
Conférences internationales	21
Tutoriaux	2
Actes et comptes-rendus de conférences ou workshops	8
Articles dans des workshops	3
Campagnes d'évaluation	4
Conférences nationales	10
Rapport d'activité	1
⌊ <i>Total des publications significatives</i>	57

Journaux internationaux

- [Reb+21a] Clément REBUFFEL, Marco ROBERTI, **LAURE SOULIER**, Geoffrey SCOUTHEETEN, Rossella CANCELLIERE et Patrick GALLINARI. “Controlling Hallucinations at Word Level in Data-to-Text Generation”. In : *Data Mining and Knowledge Discovery* (2021), accepted-in publication. URL : <https://arxiv.org/abs/2102.02810>. **Impact factor (2020) : 2.629**
- [Tam+19] Lynda TAMINE, **LAURE SOULIER**, Gia-Hung NGUYEN et Nathalie SOUF. “Offline versus Online Representation Learning of Documents Using External Knowledge”. In : *ACM Trans. Inf. Syst.* 37.4 (2019), 42 :1-42 :34. URL : <https://doi.org/10.1145/3349527>. **Impact factor (2019) : 4.676**
- [Lau+17] **LAURE SOULIER** et Lynda TAMINE. “On the Collaboration Support in Information Retrieval”. In : *ACM Comput. Surv.* 50.4 (2017), 51 :1-51 :34. **Impact factor (2019) : 7.990**
- [Lau+16d] **LAURE SOULIER**, Lynda TAMINE et Chirag SHAH. “MineRank : Leveraging users’ latent roles for unsupervised collaborative information retrieval”. In : *Inf. Process. Manage.* 52.6 (2016), p. 1122-1141. **Impact factor (2016) : 2.391**
- [Lau+14b] **LAURE SOULIER**, Lynda TAMINE et Wahiba BAHOUN. “On domain expertise-based roles in collaborative information retrieval”. In : *Inf. Process. Manage.* 50.5 (2014), p. 752-774. **Impact factor (2014) : 1.265**
- [Lau+13a] **LAURE SOULIER**, Lamjed Ben JABEUR, Lynda TAMINE et Wahiba BAHOUN. “On ranking relevant entities in heterogeneous networks using a language-based model”. In : *JASIST* 64.3 (2013), p. 500-515. **Impact factor (2013) : 2.230**

Journaux nationaux

- [Ngu+20] Gia-Hung NGUYEN, Lynda TAMINE, **LAURE SOULIER**, Clément BOSC et Nathalie SOU. “Modèle neuronal tripartite pour la représentation de documents.” In : *Document Numérique* (2020).
- [Lau+15b] **LAURE SOULIER** et Lynda TAMINE. “Vers l’intégration des métarôles pour la recherche d’information collaborative”. In : *Document Numérique* 18.2-3 (2015), p. 9-38.

2. <http://www.core.edu.au/>

Conférences internationales

- [Dje+22] Hanane DJEDDAL, Thomas GERALD, Laure SOULIER, Karen PINEL-SAUVAGNAT et Lynda TAMINE. “Does Structure Matter? Leveraging Data-to-Text Generation for Answering Complex Information Needs”. In : *ECIR*. 2022. **core ranking : A**
- [Gay+22] Jean-Baptiste GAYA, Laure SOULIER et Ludovic DENOYER. “Learning a subspace of policies for online adaptation in Reinforcement Learning”. In : *ICLR*. 2022. **core ranking : A***
- [Ger+22] Thomas GERALD et Laure SOULIER. “Continual Learning of Long Topic Sequences in Neural Information Retrieval”. In : *ECIR*. 2022. **core ranking : A**
- [Lov+21b] Jesús LOVÓN-MELGAREJO, **LAURE SOULIER**, Karen PINEL-SAUVAGNAT et Lynda TAMINE. “Studying Catastrophic Forgetting in Neural Ranking Models”. In : *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021*. Springer, 2021, p. 375-390. URL : <https://arxiv.org/abs/2101.06984>. **core ranking : A**
- [Reb+21b] Clément REBUFFEL, Thomas SCIALOM, **LAURE SOULIER**, Benjamin PIWOWARSKI, Sylvain LAMPRIER, Jacopo STAIANO, Geoffrey SCOUTHEETEN et Patrick GALLINARI. “Data-QuestEval : A Referenceless Metric for Data to Text Semantic Evaluation”. In : *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, 2021. URL : <https://arxiv.org/abs/2104.07555>. **core ranking : A**
- [Reb+20a] Clément REBUFFEL, **LAURE SOULIER**, Geoffrey SCOUTHEETEN et Patrick GALLINARI. “A Hierarchical Model for Data-to-Text Generation”. In : *ECIR*. 2020. URL : <https://arxiv.org/abs/1912.10011>. **core ranking : A**
- [Reb+20c] Clément REBUFFEL, **LAURE SOULIER**, Geoffrey SCOUTHEETEN et Patrick GALLINARI. “PARENTing via Model-Agnostic Reinforcement Learning to Correct Pathological Behaviors in Data-to-Text Generation”. In : *Proceedings of the 13th International Conference on Natural Language Generation, INLG 2020*. Association for Computational Linguistics, 2020, p. 120-130. URL : <https://arxiv.org/abs/2010.10866>. **core ranking : B**
- [Bor+19a] Patrick BORDES, Eloi ZABLOCKI, **LAURE SOULIER**, Benjamin PIWOWARSKI et Patrick GALLINARI. “Incorporating Visual Semantics into Sentence Representations within a Grounded Space”. In : *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, 2019, p. 696-707. URL : <https://doi.org/10.18653/v1/D19-1064>. **core ranking : A**
- [Zab+19] Eloi ZABLOCKI, Patrick BORDES, **LAURE SOULIER**, Benjamin PIWOWARSKI et Patrick GALLINARI. “Context-Aware Zero-Shot Learning for Object Recognition”. In : *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*. Proceedings of Machine Learning Research. PMLR, 2019, p. 7292-7303. URL : <http://proceedings.mlr.press/v97/zablocki19a.html>. **core ranking : A***
- [Car+18b] Micael CARVALHO, Rémi CADÈNE, David PICARD, **LAURE SOULIER**, Nicolas THOME et Matthieu CORD. “Cross-Modal Retrieval in the Cooking Context : Learning Semantic Text-Image Embeddings”. In : *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR*. 2018, p. 35-44. **core ranking : A***
- [Ngu+18a] Gia-Hung NGUYEN, Lynda TAMINE, **LAURE SOULIER** et Nathalie SOUF. “A Tri-Partite Neural Document Language Model for Semantic Information Retrieval”. In : *The Semantic Web - 15th International Conference, ESWC*. 2018, p. 445-461. **core ranking : A**
- [Zab+18] Eloi ZABLOCKI, Benjamin PIWOWARSKI, **LAURE SOULIER** et Patrick GALLINARI. “Learning Multi-Modal Word Representation Grounded in Visual Context”. In : *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. 2018. **core ranking : A***
- [Ngu+17a] Gia-Hung NGUYEN, **LAURE SOULIER**, Lynda TAMINE et Nathalie BRICON-SOUF. “DSRIM : A Deep Neural Information Retrieval Model Enhanced by a Knowledge Resource Driven Representation of Documents”. In : *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR*. 2017, p. 19-26. URL : <https://hal.archives-ouvertes.fr/hal-01549543>.

- [Ngu+17c] Gia-Hung NGUYEN, Lynda TAMINE, **LAURE SOULIER** et Nathalie SOUF. “Learning Concept-Driven Document Embeddings for Medical Information Search”. In : *Artificial Intelligence in Medicine - 16th Conference on Artificial Intelligence in Medicine, AIME*. 2017, p. 160-170. **core ranking : B**
- [Jab+16b] Lamjed Ben JABEUR, **LAURE SOULIER**, Lynda TAMINE et Paul MOUSSET. “A Product Feature-Based User-Centric Ranking Model for E-Commerce Search”. In : *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 7th International Conference of the CLEF Association, CLEF*. 2016, p. 174-186.
- [Lau+16b] **LAURE SOULIER**, Lynda TAMINE et Gia-Hung NGUYEN. “Answering Twitter Questions : a Model for Recommending Answerers through Social Collaboration”. In : *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016*. 2016, p. 267-276. **core ranking : A**
- [Tam+16c] Lynda TAMINE, **LAURE SOULIER**, Lamjed Ben JABEUR, Frédéric AMBLARD, Chihab HANACHI, Gilles HUBERT et Camille ROTH. “Social Media-Based Collaborative Information Access : Analysis of Online Crisis-Related Twitter Conversations”. In : *Proceedings of the 27th ACM Conference on Hypertext and Social Media, HT*. 2016, p. 159-168.
- [Tam+15] Lynda TAMINE et **LAURE SOULIER**. “Understanding the Impact of the Role Factor in Collaborative Information Retrieval”. In : *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM*. 2015, p. 43-52. **core ranking : A**
- [Lau+14a] **LAURE SOULIER**, Chirag SHAH et Lynda TAMINE. “User-driven system-mediated collaborative information retrieval”. In : *The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2014, p. 485-494. **core ranking : A***
- [Lau+13b] **LAURE SOULIER**, Lynda TAMINE et Wahiba BAHOUN. “A Collaborative Document Ranking Model for a Multi-faceted Search”. In : *Information Retrieval Technology - 9th Asia Information Retrieval Societies Conference, AIRS*. 2013, p. 109-120. **core ranking : C**
- [Lau+12a] **LAURE SOULIER**, Lamjed Ben JABEUR, Lynda TAMINE et Wahiba BAHOUN. “BibRank : a language-based model for co-ranking entities in bibliographic networks”. In : *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL*. 2012, p. 61-70. **core ranking : A**

Tutoriaux dans des conférences internationales

- [Tam+16a] Lynda TAMINE et **LAURE SOULIER**. “Collaborative Information Retrieval : Concepts, Models and Evaluation”. In : *Advances in Information Retrieval - 38th European Conference on IR Research, ECIR*. 2016, p. 885-888.
- [Tam+16b] Lynda TAMINE et **LAURE SOULIER**. “Collaborative Information Retrieval : Frameworks, Theoretical Models, and Emerging Topics”. In : *Proceedings of the 2016 ACM on International Conference on the Theory of Information Retrieval, ICTIR*. 2016, p. 7-8.

Actes et comptes-rendus de conférences ou de workshops

- [Bel+19] Patrice BELLOT et **LAURE SOULIER**, éd. *Conférence en Recherche d'Informations et Applications - CO-RIA 2019, 16th French Information Retrieval Conference. Lyon, France, May 25-29, 2019. Proceedings*. ARIA, 2019. ISBN : 978-2-9556434-2-6.
- [Bel+18] Patrice BELLOT, Chiraz TRABELSI, Josiane MOTHE, Fionn MURTAGH, Jian-Yun NIE, **LAURE SOULIER**, Eric SANJUAN, Linda CAPPELLATO et Nicola FERRO, éd. *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 9th International Conference of the CLEF Association, CLEF 2018*. T. 11018. Lecture Notes in Computer Science. Springer, 2018.
- [Cap+18] Linda CAPPELLATO, Nicola FERRO, Jian-Yun NIE et **LAURE SOULIER**, éd. *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*. T. 2125. CEUR Workshop Proceedings. CEUR-WS.org, 2018.

- [Azz+17a] Leif AZZOPARDI, Jeremy PICKENS, Chirag SHAH, **LAURE SOULIER** et Lynda TAMINE. “Report on the Second International Workshop on the Evaluation on Collaborative Information Seeking and Retrieval (ECol’2017@CHIIR)”. In : *SIGIR Forum* 51.3 (2017), p. 122-127.
- [Azz+17b] Leif AZZOPARDI, Jeremy PICKENS, Chirag SHAH, **LAURE SOULIER** et Lynda TAMINE. “Second International Workshop On the Evaluation of Collaborative Information Seeking and Retrieval (Ecol’17)”. In : *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, CHIIR*. 2017, p. 429-431.
- [Lau+16c] **LAURE SOULIER**, Lynda TAMINE, Tetsuya SAKAI, Leif AZZOPARDI et Jeremy PICKENS. “Report on the First International Workshop on the Evaluation on Collaborative Information Seeking and Retrieval (ECol’2015)”. In : *SIGIR Forum* 50.1 (2016), p. 42-48.
- [Azz+15a] Leif AZZOPARDI, Jeremy PICKENS, Tetsuya SAKAI, **LAURE SOULIER** et Lynda TAMINE. “ECol 2015 : First international workshop on the Evaluation on Collaborative Information Seeking and Retrieval”. In : *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM*. 2015, p. 1943-1944.
- [Azz+15b] Leif AZZOPARDI, Jeremy PICKENS, Tetsuya SAKAI, **LAURE SOULIER** et Lynda TAMINE, éd. *Proceedings of the 2015 Workshop on Evaluation on Collaborative Information Retrieval and Seeking, ECol*. ACM, 2015.

Articles dans des workshops internationaux avec comité de lecture

- [Ais+18] Wafa AISSA, **LAURE SOULIER** et Ludovic DENOYER. “A Reinforcement Learning-driven Translation Model for Search-Oriented Conversational Systems”. In : *Proceedings of the 2nd International Workshop on Search-Oriented Conversational AI, SCAI@EMNLP 2018, Brussels, Belgium, October 31, 2018*. 2018, p. 33-39. URL : <https://arxiv.org/abs/1809.01495>.
- [Car+18a] Micael CARVALHO, Rémi CADÈNE, David PICARD, **LAURE SOULIER** et Matthieu CORD. “Images and Recipes : Retrieval in the Cooking Context”. In : *34th IEEE International Conference on Data Engineering Workshops, ICDE Workshops 2018*. 2018, p. 169-174.
- [Ngu+16] Gia-Hung NGUYEN, Lynda TAMINE, **LAURE SOULIER** et Nathalie BRICON-SOUF. “Toward a Deep Neural Approach for Knowledge-Based IR”. In : *CoRR* abs/1606.07211 (2016). URL : <http://arxiv.org/abs/1606.07211>.

Campagnes d’évaluation

- [Zab+17] Eloi ZABLOCKI, Patrick BORDES, **LAURE SOULIER**, Benjamin PIWOWARSKI et Patrick GALLINARI. “LIP6@CLEF2017 : Multi-Modal Spatial Role Labeling using Word Embeddings”. In : *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*. 2017.
- [Jab+16a] Lamjed Ben JABEUR, **LAURE SOULIER**, Paul MOUSSET et Lynda TAMINE. “Living Ranking : from Online to Real-time Information Retrieval Evaluation”. In : *Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum*. 2016, p. 587-588.
- [Che+15] Abdelhamid CHELLAL, Lamjed Ben JABEUR, **LAURE SOULIER**, Bilel MOULAHI, Thomas PALMER, Mohand BOUGHANEM, Karen PINEL-SAUVAGNAT, Lynda TAMINE et Gilles HUBERT. “IRIT at TREC Microblog 2015”. In : *Proceedings of The Twenty-Fourth Text REtrieval Conference, TREC*. 2015.
- [Jab+15] Lamjed Ben JABEUR, **LAURE SOULIER** et Lynda TAMINE. “IRIT at CLEF 2015 : A Product Search Model for Head Queries”. In : *Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum*. 2015.

Conférences nationales

- [Erb+21] Pierre ERBACHER et **LAURE SOULIER**. “État de l’art des approches de modélisation et de simulation utilisateur pour la recherche d’information conversationnelle”. In : *CORIA 2021 - Conférence en Recherche d’Informations et Applications*. 2021.
- [Lov+21a] Jesús LOVÓN-MELGAREJO, **LAURE SOULIER**, Karen PINEL-SAUVAGNAT et Lynda TAMINE. “Oubli catastrophique et approches neuronales pour la Recherche d’Information”. In : *CORIA 2021 - Conférence en Recherche d’Informations et Applications*. 2021.
- [Luk+21] Thomas LUKA, **LAURE SOULIER** et David PICARD. “Apprentissage non supervisé de représentations de mots à l’aide de réseaux de convolution bilinéaires sur des caractères”. In : *CORIA 2021 - Conférence en Recherche d’Informations et Applications*. 2021.
- [Reb+20b] Clément REBUFFEL, **LAURE SOULIER**, Geoffrey SCOUTHEETEN et Patrick GALLINARI. “Capturing Entity Hierarchy in Data-to-Text Generative Models”. In : *Proceedings of the Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020)*. 2020.
- [Reb19] Clément REBUFFEL. “Data-to-Text : Vers la génération de texte à partir de données non-linguistiques”. In : *CONFérence en Recherche d’Informations et Applications - CORIA 2019*. Sous la dir. de Patrice BELLOT et **LAURE SOULIER**. ARIA, 2019. URL : http://www.asso-aria.org/coria/2019/RJCRI%5C_2019%5C_paper%5C_7.pdf.
- [Ngu+18b] Gia-Hung NGUYEN, Lynda TAMINE, **LAURE SOULIER** et Nathalie SOUF. “Modèle neuronal tripartite pour la représentation de documents”. In : *CONFérence en Recherche d’Informations et Applications - CORIA 2018*. 2018.
- [Ngu+17b] Gia-Hung NGUYEN, Lynda TAMINE, **LAURE SOULIER** et Nathalie BRICON-SOUF. “Modèle Neuronal de Recherche d’Information Augmenté par une Ressource Sémantique”. In : *CONFérence en Recherche d’Informations et Applications - CORIA 2017, 14th French Information Retrieval Conference*. 2017, p. 265-280.
- [Lau+16a] **LAURE SOULIER**, Lamjed Ben JABEUR, Paul MOUSSET et Lynda TAMINE. “Quels facteurs de pertinence pour la recherche de produits ecommerce?” In : *CORIA 2016 - Conférence en Recherche d’Informations et Applications- 13th French Information Retrieval Conference. CIFED 2016 Colloque International Francophone sur l’Ecrit et le Document*. 2016, p. 415-430.
- [Lau+15a] **LAURE SOULIER** et Lynda TAMINE. “Définition et exploitation des méta-rôles des utilisateurs pour la recherche d’information collaborative”. In : *CORIA 2015 - Conférence en Recherche d’Informations et Applications - 12th French Information Retrieval Conference*. 2015, p. 107-122.
- [Lau+12b] **LAURE SOULIER**, Lamjed Ben JABEUR, Lynda TAMINE et Wahiba BAHOUN. “Modèle de langue pour l’ordonnancement conjoint d’entités pertinentes dans un réseau d’informations hétérogènes”. In : *Actes du XXXème Congrès INFORSID*. 2012, p. 256-272.

Rapports d’activité

- [Ovi+20] Sharon L. OVIATT et **LAURE SOULIER**. “Conversational Search for Learning Technologies”. In : *CoRR* abs/2001.02912 (2020). arXiv : 2001.02912. URL : <http://arxiv.org/abs/2001.02912>.

Distinctions et Prix

-
- 2021 Prix du meilleur article de la *CONFérence en Recherche d’Information et Applications (CORIA)* [Luk+21]
- 2018 Prix du meilleur article du workshop sur les systèmes conversationnels à EMNLP *SCAI@EMNLP2018* [Ais+18]

- 2015 Prix du meilleur article de la *CONFérence en Recherche d'Information et Applications (CORIA)* [Lau+15a]
- 2014 Bourse ELIAS pour l'organisation d'un workshop sur la recherche d'information collaborative (ECol 2015 [Lau+16c] et 2017 [Azz+17b]). Montant : 10 000 €
- 2013 Prix du meilleur article de la conférence internationale *Asia Information Retrieval Societies (AIRS)* [Lau+13b]

Pilotage et participation à des projets ou programmes de recherche

- 2022 – 2026 *Nom* : Apprentissage Contrefactuel pour Data-to-text Contrôlé (ACDC)
Description : ANR PRCE (AAPG 2021 - CES 23)
Porteur : LIP6 (S. Lamprier)
Partenaires : LIP6 ; LAMSADE ; MNHN ; Recital
- 2018 – 2023 *Nom* : SEarch-oriented ConverSational systeMS (SESAMS)
Description : ANR JCJC (AAPG 2018 - CES 23)
***Porteuse* : LIP6 (L. Soulier)**
Partenaires : LIP6 ; CRIStAL/Inria, DIRO (Montréal)
Actions : Montage du projet, discussion avec les partenaires, rédaction de la proposition. .
Publications : [Ais+18 ; Erb+21 ; Lov+21b]
 Aucune décharge d'enseignement n'a été demandée dans le cadre de cette ANR.
- 2015 – 2018 *Nom* : MULTimodal processing of Spatial and TEMPoral expREssions (MUSTER)
Description : Chist-ERA 2014 : Human Language Understanding : Grounding Language Learning (HLU)
Porteur : LIP6 (B. Piwowarski)
Partenaires : LIP6 ; KU Leuven (Belgique), ETH Zurich (Suisse), UPV/EHU (Spain)
Participation : Encadrement des doctorants (2), Participations aux réunions du consortium et réunion Chist-ERA.
Publications : [Bor+19a ; Zab+19 ; Bor+19b ; Zab+18 ; Zab+17]
- 2014 – 2015 *Nom* : EXPloration sur l'usage des médias sociaux pour un Accès Collaboratif à l'information (EXPAC)
Appel : PEPS CNRS transdisciplinaire, HuMaIn : Humanités, Mathématiques, Sciences de l'Information
Porteur : IRIT (L. Tamine-Lechani)
Partenaires : IRIT- SIG ; CNRS March Bloch, Berlin ; Maths, SMMA, Université Paris Sorbonne
Participation : Définition du contexte et des questions de recherche, Proposition de la méthodologie scientifique, Encadrement des stagiaires (2), Participations et présentation des résultats aux réunions.
Publications : [Tam+16c ; Lau+16b]

Organisation de conférences, congrès, workshops

- 2019 SIGIR : Responsable de l'organisation locale (*local chair*).
 RJCRI@CORIA : Présidente du comité de programme (*chair*).
- 2018 CLEF : Responsable des laboratoires d'expérimentation (*lab chair*) [Bel+18 ; Cap+18].
 ECIR : Responsable de la communication (*publicity chair*).
 DECOR@ICDE : Organisatrice du workshop "Data Engineering meets Intelligent Food and COoking Recipe" (*co-chair*).

- 2017 ECol@CHIIR : Co-organisatrice de la seconde édition du workshop "Evaluation on Collaborative Information Seeking and Retrieval" (*co-chair*) [Azz+17b; Azz+17a].
- 2015 ECol@CIKM : Co-organisatrice de la première édition du workshop "Evaluation on Collaborative Information Seeking and Retrieval" (*co-chair*) [Lau+16c; Azz+15a]. Financement par une bourse ELIAS m'a été attribuée.
CLEF : Responsable de la communication (*publicity chair*).
- Avant 2014 EGC'13 et ETFA'11 : Aide logistique.

Membre de comités scientifiques

- 2018-2020 Membre du comité de direction (*Steering committee*) de CLEF (Conference and Labs of the Evaluation Forum : <http://www.clef-initiative.eu/>)

Participation à des comités de lecture

– Relectrice de journaux internationaux

- 2020 TAFFC · ACM TOIS
- 2019 ESWA
- 2018 TPAMI · TAFFC
- 2017 JASIST · IRJ
- 2016 JASIST · IJDL
- 2015 IP&M
- 2014 IP&M · ESWA

– Membre de comités de programme de conférences

- | | International | National |
|------|-------------------------------------|-----------------|
| 2022 | ECIR · SIGIR · CIKM | CIRCLE |
| 2021 | ECIR · SIGIR · CIKM · EMNLP | CORIA |
| 2020 | ECIR · SIGIR · ECAI · AAAI · COLING | CORIA (CIRCLE) |
| 2019 | ECIR · IJDL | CORIA |
| 2018 | SIGIR · WWW · ICTIR · ECIR · JCDL | CORIA |
| 2017 | SIGIR · CIKM · ICTIR | CORIA · MARAMI |
| 2016 | SIGIR · COLING · | CORIA · MARAMI |
| 2015 | SIGIR · JCDL | RJCRI |
| 2014 | AIRS | |

Conférences invitées

- 2022 - Séminaire TLC - LISN (Février 2022) : "Data-to-text generation : let speak your data fluently!"
- 2021 - Séminaire DGA (Septembre 2021) : "Deep learning pour la recherche d'information : état des lieux et perspectives"
- NaverLabs (juillet 2021) : "Data-to-text generation : let speak your data fluently!"
- Ecole d'été ETAL (juin 2021) : Recherche d'information
- 2020 - Journée "THL et multimodalité" - THL/AFIA (oct 2020) : "From multimodal representation learning to multimodal information access"
- Séminaire Marseille - équipe TALEP (octobre 2020) : "From multimodal representation learning to multimodal information access"

- 2019
- PhisIA seminar - Univ Paris Diderot (nov 2019) : "Le symbolique au service du connexionnisme et vice-versa : apprentissage de représentation augmenté, extraction d'information et bases de connaissances"
 - ERIC lab seminar (oct 2019) : "Apprentissage de représentations textuelles augmentées bases de connaissances : application à la Recherche d'information"
 - GDR IA - 2019 : "Ancrage visuel et conceptuel du texte pour l'apprentissage de représentation"
 - Panelist for the Pré-GDR TAL (March 2019)
- 2016
- Laboratoire ERIC "De la Recherche d'information collaborative à la recherche d'information socio-collaborative : fondements, modèles et perspectives"

Séjours dans des laboratoires étrangers

6-13 Septembre 2021

Chercheur invitée

Département d'informatique, Université de Turin - Italie

Titre : Data-to-text generation

Hôte : Rossella Cancelliere

Janvier 2015

Stage de Recherche Post-Doctoral

Département de mathématiques/NAXYS, Université de Namur - Belgique

Titre : Recherche d'information collaborative dans les réseaux sociaux

Encadrement : Renaud Lambiotte (Professeur Assistant, Université de Namur)

Mai - Juillet 2013

Stage de Recherche Doctoral

SC&I - Rutgers University, New Brunswick NJ - USA

Titre : Contribution à l'optimisation de sessions de recherche d'information collaborative basée sur le rôle des utilisateurs

Encadrement : Chirag Shah (Professeur Assistant, Université de Rutgers)

Publications : SIGIR 2014, IP&M 2016

Synthèse des activités pédagogiques - Sorbonne Université

Je présente dans le tableau ci-dessous les temps de service liés à l'enseignement depuis 2015, date de mon installation à l'UPMC/Sorbonne Université. Les heures reportées sont exprimées en présentiel (1hCM = 1.5 hTD).

Mes enseignements sont répartis aussi bien en licence qu'en master. Je suis **responsable de plusieurs UEs, aussi bien en licence qu'en master**.

Niveau	Année	UE	hCM	hTD	hTME	hG	Responsabilités
L1	2015-2017	Elément de programmation 1		71.75h			
L2 Info.	2015-2016	Orientation et Insertion Professionnelle		22h			Rédaction d'un sujet de projet et participation à la rédaction du sujet de ratrapage.
	2017-2020	Eléments de programmation par objets avec Java		113.75h		3	
L2 ATIC	2015-2018	Insertion Professionnelle par le stage		76h			Responsable UE - création des contenus et suivi des étudiants Responsable UE - suivi des stages et soutenances
	2015-2018	Stage en entreprise		68h			
L3 Info.	2015-2021	Technologies du Web	96h	214h		18h	Responsable UE depuis 2016 - Gestion des évaluations de projets (contrôle continu) et rédactions des sujets d'examens
M1 DAC	2015-2021	Projet logiciel DAC		54h			Proposition d'un sujet de projet et suivi d'un groupe d'étudiants Responsable UE depuis 2017, refonte des supports de cours, TD et TME, rédactions du sujet de projet et des sujets d'examens Co-responsable UE , réalisation des supports de cours/TD/TME pour la partie recherche d'information
	2016-2021	Business Intelligence	62h		72h		
	2018-2021	Recherche d'information et NLP	40h	34h	34h		
M2 DAC	2020-2021	Machine learning avancé et Deep learning		28			Co-responsable UE , réalisation des supports de cours/TD/TME pour la partie méthodologie de recherche Création d'une nouvelle partie du cours sur la méthodologie de formalisation et d'évaluation de modèles
	2019-2021	Data Science et méthodologie de recherche	32	22			
	2016-2018	Fouille de Données et Médias Sociaux	12h		12h		
	2018-2019	Recherche d'information et moteur de recherche	14h				
	2019-2021	Stage M2 DAC		20			
		Total	256h	723.5h	118h	21h	

Activités de responsabilités administratives locales et nationales

2020-2021	Présidente du comité ANR CES 38.
2019-2020	Vice-Présidente du comité ANR CES 38.
2018-2019	Membre du comité ANR CES 38.
2017-2021	Membre des comités de sélections pour les postes Maître de conférences : <ul style="list-style-type: none"> — MCF section 06 : Sorbonne Université (2018), Université Bretagne Sud (2019), IUT Toulouse (2019), CNAM (2020), INSA Lyon (2020)). — "MCF section 27 Polytech" : Université Aix-Marseille (2018) — "MCF section 65" : Sorbonne Université (2021)
2016 - 2020	Membre du comité d'organisation du Colloquium d'Informatique de Sorbonne Université. Organisation des colloquiums de Léon Bottou (Facebook), Eric Horvitz (Microsoft Research), Caroline Chavier (WIMLDS Paris).

- 2015 – 2017* **Expert Jeune docteur** pour les évaluations HCERES des écoles doctorales "Sciences Technologie Santé" (UPJV - 30 novembre 2016), "Sciences pour l'ingénieur" (UTC - 1 décembre 2016), "Sciences et ingénierie des systèmes, mathématiques, informatique" (ComUE - 7 février 2017), "Sciences et ingénierie des matériaux, mécanique, énergétique" (ComUE - 8 février 2017) et "EUCLIDE - ED multidisciplinaire" (La Rochelle, 19 octobre 2017).
- 2012 – 2014* **Représentante des doctorants** de l'Ecole Doctorale Mathématiques, Informatique, Télécommunications de Toulouse.

Participation à des jurys de thèse

- 8 octobre 2021* Examinatrice de la thèse de Hicham El Boukkouri - Université Paris Saclay. Encadrée par Pierre Zweigenbaum, Olivier Ferret et Thomas Lavergne.
- 5 février 2021* Examinatrice de la thèse de Jebril Frej - Université Grenoble Alpes. Encadrée par Jean-Pierre Chevallet et Didier Schwab.
- 8 octobre 2020* Examinatrice de la thèse de Warith Harchaoui - Université Paris Descartes. Encadrée par Charles Bouveyron.
- 4 mai 2020* Examinatrice de la thèse de Julien Tissier - Université Jean Monnet - Saint Etienne. Encadrée par Christophe Gravier et Amaury Habrard.
- 5 juillet 2019* Examinatrice de la thèse de Hady El Sahar - Université Jean Monnet - Saint Etienne. Encadrée par Christophe Gravier et Florence Laforest.
- 20 mai 2019* Examinatrice de la thèse de Hedi Ben Younes - Sorbonne Université. Encadrée par Matthieu Cord et Nicolas Thome.
- 11 octobre 2018* Examinatrice de la thèse de Micael Carvalho - Sorbonne Université. Encadrée par Matthieu Cord et Nicolas Thome.