

Romain Deveaud

romain.deveaud@gmail.com — +33 619 222 028 — Pionnat, France — linkedin.com/in/romain-deveaud/
— romaindeveaud.github.io

Description

At the intersection between an academic researcher and a technical and strategic leader, I have more than 10 years of experience in Information Retrieval, Natural Language Processing, and AI in general. I currently serve as a Manager and Tech Lead of the Search team at Qwant, a Web search engine.

I hold a PhD in Computer Science, where I've been working on modelling the implicit concepts of an ambiguous search engine query and building improved probabilistic ranking models. I authored more than 20 papers published in top tier journals and conferences related to Information Retrieval and NLP (ACM TOIS, CIKM, SIGIR, ACL), and participated several times in data challenges and workshops mainly related to Information Retrieval topics (TREC, INEX, NTCIR). I live in the French countryside with my family, advocating for working remotely in a calm and healthy environment.

Work Experience

Senior Tech Lead & Scientific Director, Qwant

January 2023 - Present

I lead and manage 6 scientists and engineers, working on R&D Search projects around Query Understanding, Document Understanding, Ranking, and Evaluation. As a Scientific Director I was responsible for defining the technical roadmap and ensuring scientific rigor across all teams.

- Designed and implemented a patented Query Performance Prediction system to select the optimal index from over 1,000 sources, achieving > 90% accuracy at < 1 ms latency.
- Led the development of an end-to-end hybrid search stack, integrating lexical and dense features (fine-tuned Matryoshka embeddings) into GBDT ranking models.
- Architected core query understanding modules, including intent classification, topic classification, spellchecking, semantic query rewriting (aliases, acronyms), and trend detection.
- Engineered a scalable ecosystem for isolated experimentation, industrialized hyperparameter optimization, and automated model retraining to prevent performance drift.

Technologies: Vespa, Apache Spark, FAISS, floret/fasttext, scikit-learn, pytorch, pandas, polars, Jax, Argo, Kubernetes, MLFlow, Gitlab CI/CD, Triton serving, huggingface ecosystem.

Research Engineer, Qwant

08/2021 - 01/2023

Full remote.

Member of the CoreSearch team.

- Built the offline ranking evaluation platform.
- Introduced and developed probabilistic click models for computing noise-free and unbiased relevance estimates of documents.
- Experimented with unbiased Learning to Rank using implicit user feedback.
- Developed a sequence labelling model (Bi-LSTM w/ CRF layer) for tagging entities and attributes within queries, enabling a more accurate display of entity cards.
- Explored and implemented a query suggestion solution (Click Graph random walk, semantic similarities)

Technologies: Vespa, pandas, pytorch, scikit-learn, huggingface ecosystem, Metabase, Apache Spark, MLFlow, Docker, Kubernetes, Argo, Gitlab CI/CD

Senior Research Engineer

12/2020 - 08/2021

Aiway

Full remote.

Romain Deveaud

Providing tailored machine learning solutions to clients.

- Natural Language Understanding of text and audio messages. Sequence labelling with distant supervision (no training data) through a contextual Bi-LSTM for tagging entities, followed by ranked lists fusion for item retrieval.
- Time-series consumption forecasting of a water distribution network system.
- Researching a Conversational Recommender System using Reinforcement Learning.

Freelance researcher

11/2019 - 12/2020

Unbiased Learning to Rank for search and recommendation. Using counterfactual reasoning to limit position biases in rankings, and popularity and positivity biases in recommendation.

Postdoc researcher

10/2018 - 10/2019

OpenEdition, CNRS, Aix-Marseille Université

Full remote.

Appropriation du Savoir Ouvert project, publicly funded. Researching and analysing web logs data in order to describe how people visit and consume online open access academic documents. Analysis of web server logs to understand how people interact and consume the content of an open access platform for academic content. Implemented predictive models of dwell time, session-based document embeddings, taxonomy of returning users...

Supervised by: *Marin Dacos* and *Patrice Bellot*.

Postdoc researcher

05/2018 - 08/2018

Laboratoire Informatique d'Avignon (Computer science lab of Avignon)

Working from home, fully remote.

AMIS project, publicly funded by the French government. Indexing and summarisation of news audio files (github.com/amis-anr/audio-summariser).

Unsupervised identification of salient and relevant audio parts from multi-lingual news reports. Development of an interface for human annotation and further supervised training based on the informativeness of the previously extracted audio segments.

Supervised by: *Eric SanJuan*.

Postdoc researcher

01/2016 - 12/2016

IRIT, Université Paul Sabatier, Toulouse

Working from home, fully remote, worked part-time (50%).

Projet ANR CAAS, *Contextual Analysis and Adaptive Search* ; Learning to Rank applied to predicting the effectiveness of document retrieval systems configurations. Reducing the combinatorial problem of finding the best parameters of an ad-hoc document retrieval system by learning to rank grids of parameter configurations according to their full retrieval performance.

Supervised by: *Josiane Mothe*, collaborating with *Jian-Yun Nie*.

Senior developer

01/2015 - 05/2015

Yellow Pages Canada, Montréal, Canada.

Team Search & Relevance.

Implemented an unsupervised categorisation of unknown words and queries using word embeddings and an in-house ontology.

Postdoc researcher

10/2013 - 11/2014

Terrier team, University of Glasgow, UK.

European project FP7 SMART, information retrieval and recommendation in the context of smart cities. Open source development (see github.com/SmartSearch).

Development of unsupervised (probabilistic) and supervised (Learning to Rank) models for the recommendation of venues and points-of-interest in a zero-query setting. Supervised by: *Iadh Ounis* and *Craig Macdonald*.

Graduate student

10/2010 - 09/2013

Laboratoire Informatique d'Avignon (Computer science lab of Avignon).

Romain Deveaud

ANR project CAAS, *Contextual Analysis and Adaptive Search*.

Modelling of the latent concepts of keyword queries by developing query-biased topic models.

Participation in several workshops. Supervised by: *Eric SanJuan* and *Patrice Bellot*.

Web developer

09/2009 - 08/2010

Direction of Information Systems, Université d'Avignon et des Pays de Vaucluse.

Developing free and open source web applications designed to help collaborative work (see github.com/UAPV).

Supervised by: *Thierry Spriet*.

Software developer

06/2009 - 08/2009

Laboratoire Informatique d'Avignon (Computer science lab of Avignon).

European project FP6 LUNA, *Spoken language understanding in multilingual communication systems*.

Developing research prototypes and interfaces.

Supervised by: *Frédéric Béchet*.

Education

Ph.D. in Computer Science

2013

Vers une représentation du contexte thématique en Recherche d'Information (Representing Topical Query Contexts in Information Retrieval), Very honorable mention.

Supervisors: *Eric SanJuan* and *Patrice Bellot*.

Viva passed at l'Université d'Avignon et des Pays de Vaucluse, November 29th, 2013.

Jury president : *Josiane Mothe*.

Reviewers : *Jian-Yun Nie* and *Philippe Mulhem*.

Examinators : *Jaap Kamps*, *Benjamin Piwowarski*, and *Jacques Savoy*.

M.Sc. in Computer Science

2010

Université d'Avignon et des Pays de Vaucluse.

Working part-time during last year, thesis: *Open-source software at the University of Avignon*.

B.Sc. in Computer Science

2008

Université d'Avignon et des Pays de Vaucluse.

DUT in Computer Science (2-year French diploma)

2007

Université de la Méditerranée, Aix-en-Provence.

Specialised in software engineering.

Technical Skills

These are just a few highlighted skills, but being a fast learner I can quickly be proficient in new technologies if needed.

Programming : Python, R, JavaScript, Java, PHP, Ruby, bash, ...

Frameworks & libraries : pandas, pytorch, scikit-learn, matplotlib, spark, huggingface transformers, jQuery...

Search engine tools : Vespa, Terrier

DevOps & teamwork : kubernetes, Argo, Hadoop, Git, CI/CD

Other : L^AT_EX, Unix, scientific communication, vulgarisation

Awards

Best paper award: CORIA 2013 [17] (the English version [3] currently has more than 330 citations)

Best paper runner-up award: MajecSTIC 2009 [22]

Publications

Journals

- [1] **Romain Deveaud**, Josiane Mothe, Md Zia Ullah and Jian-Yun Nie. Learning to Adaptively Rank Document Retrieval System Configurations. *ACM Transactions on Information Systems (TOIS)*, Volume 37 Issue 1, January 2019.
- [2] **Romain Deveaud** and Florian Boudin. De quoi parle ce Tweet? Resumer Wikipédia pour contextualiser des microblogs. *Interaction Intelligence Information*, 2014.
- [3] **Romain Deveaud**, Eric SanJuan, and Patrice Bellot. Accurate and Effective Latent Concept Modeling for Ad Hoc Information Retrieval. *Document Numérique*, 2014.

Conferences

- [4] Carlos-Emiliano González-Gallardo, **Romain Deveaud**, Eric SanJuan, and Juan-Manuel Torres-Moreno. Audio Summarization with Audio Features and Probability Distribution Divergence. *Proceedings of CICLing*, 2019.
- [5] **Romain Deveaud**, Josiane Mothe and Jian-Yun Nie. Learning to Rank System Configurations. *Proceedings of CIKM*, pages 2001–2004, 2016. Acceptance rate: 23%.
- [6] **Romain Deveaud**, Véronique Moriceau, Josiane Mothe and Eric SanJuan. Informativeness for Adhoc IR evaluation: a measure that prevents assessing individual documents In *Proceedings of ECIR*, pages 818–823, 2016. Acceptance rate: 25%.
- [7] **Romain Deveaud**, M-Dyaa Albakour, Craig Macdonald and Iadh Ounis. Experiments with a Venue-Centric Model for Personalised and Time-Aware Venue Suggestion In *Proceedings of CIKM*, pages 53–62, 2015. Acceptance rate: 16%.
- [8] Graham McDonald, **Romain Deveaud**, Richard McCreddie, Craig Macdonald and Iadh Ounis. Tweet Enrichment for Effective Dimensions Classification in Online Reputation Management. In *Proceedings of ICWSM*, pages 654–657, 2015. Acceptance rate: 27%.
- [9] Thibaut Thonet, **Romain Deveaud**, Craig Macdonald and Iadh Ounis. Suggestion contextuelle composite. In *Actes de CORIA*, pages 89–104, 2015. Acceptance rate: 24%.
- [10] **Romain Deveaud**, M-Dyaa Albakour, Craig Macdonald and Iadh Ounis. On the Importance of Venue-Dependent Features for Learning to Rank Contextual Suggestions In *Proceedings of CIKM*, pages 1827–1830, 2014. Acceptance rate: 25%.
- [11] **Romain Deveaud**, M-Dyaa Albakour, Jarana Manotumruksa, Craig Macdonald and Iadh Ounis. Diversifying Contextual Suggestions from Location-based Social Networks In *Proceedings of CIKM (demo)*, pages 125–134, 2014. Acceptance rate: 39%.
- [12] M-Dyaa Albakour, **Romain Deveaud**, Craig Macdonald and Iadh Ounis. Diversifying Contextual Suggestions from Location-based Social Networks In *Proceedings of IIRX*, pages 125–134, 2014. Acceptance rate: 46%.
- [13] **Romain Deveaud**, Eric SanJuan, and Patrice Bellot. Are Semantically Coherent Topic Models Useful for Information Retrieval? In *Proceedings of ACL*, pages 148–152, 2013. Acceptance rate: 24%.
- [14] **Romain Deveaud**, Eric SanJuan, and Patrice Bellot. Estimating Topical Context by Diverging from External Resources. In *Proceedings of SIGIR*, pages 1001–1004, 2013. Acceptance rate: 35%.
- [15] **Romain Deveaud**, Eric SanJuan, and Patrice Bellot. Unsupervised Latent Concept Modeling to Identify Query Facets. In *Proceedings of OAIR*, pages 93–100, 2013. Acceptance rate: 22.5%.
- [16] Ludovic Bonnefoy, Vincent Bouvier, **Romain Deveaud**, and Patrice Bellot. Vers une détection en temps réel de documents web centrés sur une entité donnée. In *Actes de CORIA*, pages 21–36, 2013. Acceptance rate: 20%.
- [17] **Romain Deveaud**, Ludovic Bonnefoy, and Patrice Bellot. Quantification et identification des concepts implicites d’une requête. In *Actes de CORIA*, pages 159–174, 2013. Acceptance rate: 20%, **Best paper award**.

- [18] **Romain Deveaud** and Florian Boudin. Contextualisation automatique de tweets à partir de wikipédia. In *Actes de CORIA*, pages 125–140, 2013. Acceptance rate: 20%.
- [19] **Romain Deveaud** and Patrice Bellot. Combinaison de ressources générales pour une contextualisation implicite de requêtes. In *Actes de TALN*, pages 479–486, 2012.
- [20] **Romain Deveaud**, Eric SanJuan, and Patrice Bellot. Ajout d’informations contextuelles issues de Wikipédia pour la recherche de passages. In *Actes de TALN*, pages 63–68, 2011.
- [21] **Romain Deveaud**, Florian Boudin, Eric SanJuan, and Patrice Bellot. Correction de césures et enrichissement de requêtes pour la recherche de livres. In *Actes de CORIA*, pages 89–96, 2011.
- [22] Ludovic Bonnefoy, **Romain Deveaud**, and Eric Charton. Interrogations de moteurs de recherche par des requêtes formulées en langage naturel. In *Actes de MajecSTIC*, 2009. “Best hope” special award.

Workshops

- [23] Richard McCreadie, **Romain Deveaud**, M-Dyaa Albakour, Stuart Mackie, Craig Macdonald, Iadh Ounis, Thibaut Thonet and Bekir Taner Dinger. University of Glasgow at TREC 2014: Experiments with Terrier in Contextual Suggestion, Temporal Summarisation and Web Tracks. In *Proceedings of TREC*, 2014.
- [24] Graham McDonald, **Romain Deveaud**, Richard McCreadie, Timothy Gollins, Craig Macdonald and Iadh Ounis. University of Glasgow Terrier Team / Project Abacá at RepLab 2014: Reputation Dimensions Task. In *Proceedings of RepLab @ CLEF 2014*, 2014.
- [25] **Romain Deveaud**, M-Dyaa Albakour, Craig Macdonald, and Iadh Ounis. Challenges in Recommending Venues within Smart Cities. In *Proceedings of the ECIR 2014 Workshop on Information Access in Smart Cities*, i-ASC, 2014.
- [26] **Romain Deveaud** and Florian Boudin. Effective Tweet Contextualization with Hashtags Performance Prediction and Multi-Document Summarization. In *Proceedings of INEX*, 2013.
- [27] **Romain Deveaud** and Eric SanJuan. LIA at the NTCIR-10 INTENT Task. In *Proceedings of NTCIR*, 2013.
- [28] **Romain Deveaud**, Eric SanJuan, and Patrice Bellot. LIA at TREC 2012 Web Track : Unsupervised Search Concepts Identification from General Sources of Information. In *Proceedings of TREC*, 2012.
- [29] Ludovic Bonnefoy, **Romain Deveaud**, and Patrice Bellot. Do Social Information Help Book Search? In *Proceedings of INEX*, 2012.
- [30] **Romain Deveaud** and Florian Boudin. LIA/LINA at the INEX 2012 Tweet Contextualization track. In *Proceedings of INEX*, 2012.
- [31] Pierre Joulin, **Romain Deveaud**, Eric SanJuan-Ibekwe, Jean-Marc Francony, and Françoise Para. Design, implementation and experiment of a YeSQL Web Crawler. In *Proceedings of the SIGIR 2012 Workshop on Open Source Information Retrieval*, OSIR, 2012.
- [32] **Romain Deveaud**, Eric SanJuan, and Patrice Bellot. Social Recommendation and External Resources for Book Search. In *Proceedings of INEX*, pages 68–79, 2011.
- [33] **Romain Deveaud**, Eric SanJuan, and Patrice Bellot. LIA at TREC 2011 Web Track: Experiments on the Combination of Online Resources. In *Proceedings of TREC*, 2011.
- [34] **Romain Deveaud**, Florian Boudin, and Patrice Bellot. LIA at INEX 2010 Book Track. In *Proceedings of INEX*, pages 118–127, 2010.

PhD thesis

- [35] **Romain Deveaud**. Vers une représentation du contexte thématique en Recherche d’Information, 2013.