Ioannis Arapakis

Curriculum Vitae

Contact Information

Carrer de Lepant 340, 3-1, 08025, Barcelona, Spain
arapakis.ioannis@gmail.com
Skype: ioannis.arapakis
+34 6179 80472
..es.linkedin.com/in/ioannisarapakis
https://iarapakis.github.io
iarapakis

Summary

- Computer Scientist with more than 8 years of experience in both academia and industry
- Background in Data Mining, Information Retrieval, Human-Computer Interaction, and Machine Learning
- Applied research on Web Search, User Engagement, and Computational Advertising

Work Experience

2016–Present **Eurecat** Barcelona, Spain
Senior Data Scientist

Main tasks:

- Performing applied science in the context of web search, user engagement, predictive modelling, social media, among others.
- Ongoing engagement with industry and clients like Privalia, La Vanguardia, Zurich Insurance, and Cinesa.
- Collaborated with Principal Research Scientist Fabrizio Silvestri at Yahoo Labs London, under the academic program ARC, on a project on computational advertising. Contributed to a solution based on information retrieval techniques to compute the relevance of an ad to a query in order to reduce the amount of ads considered not relevant, resulting in an increase of 6% in CTR.

2011–2016 **Yahoo Labs** Barcelona, Spain
Research Scientist

Main tasks:

- Worked under three different groups: (1) User Engagement, (2) Web Retrieval, and (3) Ad Processing and Retrieval.
- Lead and executed novel research projects on web search **[5,6,8,13,17,20,23]**, cold-start recommendation **[1]**, user engagement **[8,13,17,18]**, ad retrieval **[6]**, and discovery of POI **[2]**, producing data-driven research insights; generated novel metrics and experiments to assess online engagement and conversion; and presented key findings and research product recommendations.
- Demonstrated scientific excellence and world-class external visibility through publications, presentations and external collaborations at top-tier international conferences, high-impact journals, and academic venues; supervised Ph.D. and M.Sc. students.
- Worked with product teams and business units to identify, define, and execute strategic research directions.
- Coordinated the preparation of several research grants with various academic and industrial partners; for example, the FP7 ICT 2013 MULTISENSOR was ranked second (Score: 14.5/15) and a collaboration under Yahoo's Faculty Research and Engagement Program (FREP) 2013 was established with the University of Edinburgh; managed budget of about 300K Euros.

Education

2007–2010 **Ph.D.** in Computer Science **University of Glasgow**
Thesis: "Affect-Based Information Retrieval"
Supervisors: Prof. Joemon M. Jose, Dr. Phil Gray, Prof. C. J. Keith van Rijsbergen
Glasgow, United Kingdom

2005–2007 **M.Sc.** in Information Technology **Royal Institute of Technology (KTH)**
with a specialisation in Interactive Systems Engineering
Supervisor: Asst. Prof. Teresa Cerratto Pargman
Stockholm, Sweden

Research Interests

- **Data Mining** (pattern recognition, predictive modelling, statistical inference, time series analysis, deep-learning)
- **Information Retrieval** (multimedia mining and search, user modelling, personalised search systems, recommender systems, evaluation and applications)
- **Human-Computer Interaction** (experimental methods, user engagement, neuro-physiological signal processing, sentiment analysis)

Research Skills

- Statistical inference (parametric/non-parametric methods) and time series analysis.
- Machine learning (supervised and unsupervised learning, feature engineering, data transformations).
- Pertinent modalities and techniques like neuro-physiological signal processing (EEG, EDA, BVP, HR), observer (eye tracking, facial expression analysis), and self-report methods (questionnaires, think-aloud protocols, interviews).
- Observational and experimental user studies (online and offline), A/B testing.
- Modern application design and development methodologies (ethno-mining, grounded theory methodology).

Technical Skills

- **Programming:** Java (expert), Python (basic), Matlab (prior experience), C++ (prior experience), UNIX shell scripting (proficient).
- **Scripting Languages, Web Tools and Frameworks:** J2EE, JavaServer Pages, JavaScript, PHP, HTML, DHTML (proficient).
- **Data Mining & ML Tools:** Weka (expert), Caret (expert), TensorFlow (basic).
- **Statistical Analysis:** R (expert), SPSS (expert).
- **DBMSs:** MySQL (prior experience).
- **Large-scale Analytics:** Hadoop/Pig (experience with data processing at Yahoo Labs grid), Spark (basic).
- **Web Retrieval and Crawling:** Indri (prior experience), Apache Solr (prior experience).
- **Experience in all project stages:** Research, design, development, testing, and debugging.

Internal Projects

- **Impact of search latency on user engagement in web search [13, 15, 17]:** The users show high variation in the way they perceive search response latency of a web service. As this depends on factors like the user demographics, context, and other, there is an opportunity to serve the search results to each user at custom latencies depending on the estimated behavioural impact, while minimising the usage of hardware resources. My focus in this project has been on (1) devising mechanisms for accurate prediction of the impact of response latency on user engagement, and (2) developing models for personalising response latency on a per-user basis, eventually aiming to achieve financial cost savings for the search engine company.
- **Ad Retrieval [6]:** In this project, I investigated methods for effective <*query*, *ad*> matching, to promote higher engagement with ads, improve CTR, and increase advertisement revenues. More specifically, I contributed in identifying latent associations between queries and ads, as well as coordinated a number of editorial studies for the creation of annotated datasets that were used to train machine-learning models.
- **Predicting long-term user engagement:** In this work, I investigated users' long-term engagement with a web site. In particular, I analysed the interactions of a large sample of users with the Yahoo News portal and characterised long-term user engagement by means of different metrics. I also explored the feasibility of predicting long-term engagement using short-term interaction signals, by employing both classification and regression techniques. The findings of this work have important implications for Internet companies, where long-term user engagement is likely to lead to better monetisation and customer retention (i.e., reduced churn).
- **Modelling news article quality [1, 3, 4, 16]:** The goal of this project is to identify certain proxies for characterising news article quality in an automatic and scalable way. In order to learn predictive models that can accurately estimate the perceived quality of news articles, I generated a ground-truth dataset with the help of expert judges. To this end, I performed an editorial study and created an in-domain, annotated news corpus. Using this annotated news corpus, I learned models that use shallow, syntactic, probabilistic and more complex features, to perform accurate, domain-specific, news article quality prediction and sentiment classification.
- **Scalable mouse tracking analysis for inferring user intent [8, 18]:** I performed a number of crowdsourcing studies for (1) observing how users respond to online news in the presence or lack of interest, and (2) predicting user engagement in sponsored search with specific SERP modules and Direct Displays (e.g., ads). I collected mouse tracking data, which are known to correlate with visual attention, and examined how cursor behaviour can inform user engagement measures by using an unsupervised learning approach. I researched methods that work even at the absence of traditional metrics like click-through rate or dwell time, and rely on implicit, low-cost, and scalable feedback signals.
- **Discovery and localisation of points of interest [2]:** This project focused on accurately locating points of interest visible in photos, by exploiting the location information and compass orientation supplied by modern photo cameras. My contribution involved a rigorous evaluation of the models' performance against different baselines, using machine learning and statistical analysis techniques.
- **Linking entities to past articles [4, 19]:** One aspect of engagement deals with keeping users on the site longer, by allowing them to navigate through content with enhanced, click-through experiences. So far these links have been manually curated by professional editors, and due to the manual effort involved, the use of such links has been limited. To address this issue, I participated in the design and evaluation of an automated approach to detecting and linking newsworthy events to associated articles.

Awards

2010 **Best Student Paper Award** **SIGIR**
A comparison of general vs personalised affective models for the prediction of topical relevance.
In *Proceedings of the 33rd Annual ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 371-378.

Publications

Journal Papers

1. **I. Arapakis**, B. Barla Cambazoglu, and M. Lalmas. **On the Feasibility of Predicting News Popularity at Cold Start**. In *Journal of the American Society for Information Science and Technology*, 2016.
2. B. Thomee, **I. Arapakis**, and D. A. Shamma. **Finding social points of interest from georeferenced and oriented online photographs**. In *ACM Transactions on Multimedia Computing*, 12(2), 2016.
3. **I. Arapakis**, M. Lalmas, B. Barla Cambazoglu, M. C. Marcos, and J. M. Jose. **User Engagement in Online News: Under the Scope of Sentiment, Interest, Affect, and Gaze**. In *Journal of the American Society for Information Science and Technology*, 65(10):1988-2005, 2014.
4. **I. Arapakis**, M. Lalmas, H. Ceylan, and P. Donmez. **Automatically Embedding Newsworthy Links to Articles: From Implementation to Evaluation**. In *Journal of the American Society for Information Science and Technology*, 65(1):129-145, 2013.
5. I. Lopatovska, and **I. Arapakis**. **Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction**. In *Information Processing & Management*, 47(4):575-592, 2011.

Conference Papers

6. L. Aiello, **I. Arapakis**, R. Baeza-Yates, X. Bai, N. Barbieri, A. Mantrach, and F. Silvestri. **The Role of Relevance in Sponsored Search**. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, 2016.
7. **I. Arapakis**, F. Peleja, B. Barla Cambazoglu, and J. Magalhães. **Linguistic Benchmarks of Online News Article Quality**. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2016.
8. **I. Arapakis**, and L. A. Leiva. **Predicting User Engagement with Direct Displays in Web Search Using Mouse Cursor Features**. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2016.
9. F. Peleja, **I. Arapakis**, and J. Magalhães. **Explanatory opinions: to whom or what is all the fuzz about?**. In *Proceedings of the 6th BCS-IRSG Symposium on Future Directions in Information Access*, pp. 74-77, 2015.
10. **I. Arapakis**, B. Barla Cambazoglu, and L. A. Leiva. **Know Your Onions: Understanding the User Experience with the Knowledge Module in Web Search**. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pp. 1695-1698, 2015.
11. A. Pereda-Baños, **I. Arapakis**, and M. Barreda-Àngeles. **On Human Information Processing in Information Retrieval** (Position Paper). In *Proceedings of the 2015 SIGIR Workshop on Neuro-Physiological Methods in IR*, 2015.
12. M.-C. Marcos, F. Gavin, and **I. Arapakis**. **Effects of Snippets on User Experience in Web Search**. In *Proceedings of the 16th International Conference on Human-Computer Interaction (Interacción)*, 2015.
13. M. Barreda-Àngeles, **I. Arapakis**, X. Bai, B. Barla Cambazoglu, and A. Pereda-Baños. **Unconscious Physiological Effects of Search Latency on Users and Their Click Behaviour**. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 203-212, 2015.
14. S. Vrochidis, I. Kompatsiaris, G. Casamayor, **I. Arapakis**, R. Busch, V. Alexiev, E. Jamin, M. Jugov, N. Heise, T. Forrellat, D. Liparas, L. Wanner, I. Miliaraki, V. Aleksic, K. Simov, A.-M. Soro, M. Eckhoff, T. Wagner, and M. Puigbo. **MULTISENSOR: Development of multimedia content integration technologies for journalism, media monitoring and international exporting decision support**. In *2015 IEEE International Conference on Multimedia & Expo*, 2015.
15. **I. Arapakis**, X. Bai, and B. Barla Cambazoglu. **A Controlled User Study on Human Perception of Web Search Latency**. In *TechPulse* (Internal Yahoo conference), 2014.
16. **I. Arapakis**, B. Barla Cambazoglu, and M. Lalmas. **On the Feasibility of Predicting News Popularity at Cold Start**. In *Proceedings of the 6th International Conference on Social Informatics*, pp. 290-299, 2014.
17. **I. Arapakis**, X. Bai, and B. Barla Cambazoglu. **Impact of Response Latency on User Behavior in Web Search**. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 103-112, 2014.
18. **I. Arapakis**, M. Lalmas, and G. Valkanas. **Understanding Within-Content Engagement through Pattern Analysis of Mouse Gestures**. In *Proceedings of the 23rd International Conference on Information and Knowledge Management*, pp. 1439-1448, 2014.
19. H. Ceylan, **I. Arapakis**, P. Donmez, M. Lalmas. **Automatically Embedding Newsworthy Links to Articles**. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pp. 1502-1506, 2012.
20. **I. Arapakis**, K. Athanasakos, and J. M. Jose. **A comparison of general vs personalised affective models for the prediction of topical relevance**. In *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 371-378, 2010.
21. **I. Arapakis**, I. Konstas, and J. M. Jose. **Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance**. In *Proceedings of the Seventeen ACM International Conference on Multimedia*, pp. 461- 470, 2009.
22. **I. Arapakis**, I. Konstas, J. M. Jose, and I. Kompatsiaris. **Modeling facial expressions and peripheral physiological signals to predict topical relevance**. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 728-729, 2009.
23. **I. Arapakis**, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose. **Enriching user profiling with affective features for the improvement of a multimodal recommender system**. In *Proceeding of the ACM International Conference on Image and Video Retrieval*, pp. 1-8, 2009.
24. **I. Arapakis**, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose. **Integrating facial expressions into user profiling for the improvement of a multimodal recommender system**. In *Proceedings of the 2009 IEEE International Conference on Multimedia and Expo*, pp. 1440-1443, 2009.
25. **I. Arapakis**, J. M. Jose, and P. D. Gray. **Affective feedback: an investigation into the role of emotions in the information seeking process**. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 395-402, 2008.

Teaching Experience

2016 **Universitat Pompeu Fabra, Graduate School of Economics** Barcelona, Spain
Professor at Master's Degree in Data Science: Data Science Program

- Data Visualization: This course deals with the visualization of large data sets as a means of communicating relevant data patterns aims in the form of graphical displays, in order to interpret and understand the data
- Workshop on Distributed Machine Learning: This workshop examines the basic concepts (RDDs, transformations, runtime architecture) behind distributed Machine Learning (ML) using Spark. In addition, it motivates the use of distributed ML and introduces MLlib, a library of machine learning functions.

2012–2016 **Universitat Pompeu Fabra** Barcelona, Spain
Interviewee for Digital Master of Science in Digital Information with a specialisation in Search Engines

- Provided a written interview that discusses the state-of-the-art in user engagement and the emerging research problems
- Interacted with the M.Sc. students through an online, distance-based learning environment
- Supported the students in their written assignment and provided expert feedback

2007–2008 **University of Glasgow** Glasgow, United Kingdom
Level 2 Tutor and Degree Exam Marker
Responsible for tutoring the following modules in CS undergraduate studies:

- Java Programming 2 (second year)
- Object-Oriented Software Engineering 2 (second year)
- Algorithms and Data Structures 2 (second year)

Supervision Experience

2014–Present **School of Information & Library Science, University of North Carolina** USA
Member of the Ph.D. committee of Ph.D. candidate Ashlee Edwards

Interns

- David Warnock (University of Glasgow, School of Computing Science), 2012
- Ahmet Iscen (Department of Computer Engineering, Bilkent University), 2014
- Ilker Sarac (Department of Computer Engineering, Bilkent University), 2014
- Shatlyk Ashyralyev (Department of Computer Engineering, Bilkent University), 2014
- Giwrgos Hompis (Computer Science Department, University Of Crete), 2015
- Filipa Peleja (Universidade Nova de Lisboa), 2015

Editorial Activities

- Reviewer for Information Processing & Management International Journal (2008 - to date)
- Reviewer for ACM Transactions on Information Systems (2016 - to date)
- Reviewer for IEEE Transactions on Multimedia Journal (2009 - to date)
- Reviewer for Journal of the Association for Information Science and Technology (2011 - to date)
- Reviewer for IEEE Internet Computing Journal (2012 - to date)
- Reviewer for Computers in Biology and Medicine Journal (2013 - to date)
- Reviewer for Frontiers Journal (2013 - to date)
- Reviewer for IEEE Transactions on Knowledge and Data Engineering (2015 - to date)
- Reviewer for SCIENCE CHINA Information Sciences (2015 - to date)
- Reviewer for Journal of Personalization Research (2016 - to date)

Program Committee Memberships

- WWW 2016
- ICDM 2016 (Sponsorship Chair)
- ICPRAM 2016-2017
- ACM SIGIR 2014-2017
- ECIR 2017 (Senior PC)
- ACM CIKM 2015
- ACM EMNLP 2015
- ASE/IEEE SocialCom 2012

- ICMCCA 2012
- Workshop on Neuro-Physiological Research in Information Retrieval, SIGIR 2015
- Workshop on Emotions and Personality in Personalized Services, UMAP 2013-2016
- Workshop "On the Role of Engagement in Human Information Interaction: From Research to Implementation" for iConference 2015
- International Workshop on Multimodal Media Data Analytics, ECAI 2016

Invited Talks

- "**User Behaviour Modelling: Online and Offline Methods, Metrics, and Challenges**", Keynote speaker at 1st International Workshop on System And User Centered Evaluation Approaches in Interactive Information Retrieval, CHIIR 2016, Chapel Hill, Carolina, USA, March 17, 2016.
- "**User Behaviour Modelling: Online and Offline Methods, Metrics, and Challenges**", Telefonica Research, Barcelona, Spain, January 21, 2016.
- "**Conscious and Unconscious Effects of Search Latency on User Behavior in Web Search**", Microsoft STC-E, Munich, Germany, June 17, 2015.
- "**User Engagement in Web Search: Response Latency, Engagement Metrics, and Scalability**", iConference 2015 Workshop, Newport Beach, California, US, March 24, 2015.
- "**System and User Aspects of Web Search Latency**", Universitat Pompeu Fabra, Department of Economics and Business, Barcelona, Spain, February 10, 2015.
- "**A Controlled User Study on Human Perception of Web Search Latency**", TechPulse 2014 (Internal Yahoo Conference), Santa Clara, CA, US, December 9, 2014.
- "**Impact of Response Latency on User Behavior in Web Search**", Telefonica Research, Barcelona, Spain, October 20, 2014.
- "**User Engagement in Online News: Under the Scope of Sentiment, Interest, Affect, and Gaze**", Universitat Pompeu Fabra, IDEC, Barcelona, Spain, February 4, 2014.
- "**The Face of Interest or the Face of Indifference? Predicting Topical Relevance Using Real-Time Facial Expression Analysis**", Polytechnic University of Valencia, Spain, February 1, 2013.
- "**Predicting News Popularity in Online Communities**", Yahoo Labs Winter Science Week, Sunnyvale, CA, US, March 7, 2013.
- "**Affective Feedback: What do user emotions tell us about topical relevance?**", Dublin City University, Dublin, Ireland, June 5, 2009.
- "**Affective Feedback: An Investigation into the Role of Emotions in the Information Seeking Process**", Centre for Research & Technology (ITI), Thessaloniki, Greece, September 1, 2008.

Languages

- **Greek:** Listening (C2), Reading (C2), Speaking (C2), Writing (C2)
- **English:** Listening (C2), Reading (C2), Speaking (C2), Writing (C2)
- **Spanish:** Listening (B1), Reading (B1), Speaking (B1), Writing (B1)

References

Dr. Fabrizio Silvestri	**Dr. B. Barla Cambazoglu**	**Prof. Joemon M. Jose, Fellow BCS, CITP**
Facebook	NTENT	University of Glasgow
10 Brock St		University Avenue
NW1 3FG, London, United Kingdom	Barcelona, Spain	G12 8QQ, Glasgow, UK
Email: fsilvestri@fb.com	**Email:** yurasin@gmail.com	**Email:** joemon.jose@glasgow.ac.uk