

FINAL REPORT

Student name: Maria Maistro

Cycle: 30°

Curriculum: Information and Communication Science and Technologies (ICT)

Supervisor name: Nicola Ferro

Thesis title (final): Exploiting User Signals and Stochastic Models to Improve Information Retrieval Systems and Evaluation

PART 1 - COURSES, CONFERENCES AND MOBILITY

Courses for Ph.D. Students

- Bayesian Machine Learning, Prof. Giorgio Maria Di Nunzio;
- Applied Functional Analysis, Prof. G. Pillonetto;
- Applied Linear Algebra, Prof. G. Picci;
- Real-Time Systems and Applications, Prof. G. Manduchi;
- Statistical Methods, Prof. L. Finesso;
- Academic English Course for Ph.D. Students, Dr. Gillian Davies.

Bachelor [LT] and Master Degree [LM] Courses

- [LT] Dati e Algoritmi 1, Prof. A. Pietracaprina;
- [LM] Basi di Dati, Prof. N. Ferro;
- [LM] Reperimento dell'Informazione, Prof. M. Agosti;
- [LM] Reti di Calcolatori, Prof. A. Zanella.

Summer schools, short courses, tutorials

- 10th European Summer School in Information Retrieval, (ESSIR2015), Thessaloniki, Greece, 31 August - 4 September 2015;
- “Statistical Significance Testing in Theory and in Practice” Tutorial, Prof. Ben Carterette, from University of Delaware, September 27, 2015;
- “Bibliographic resources and research tools for Industrial Engineering”, Library short course, January 21, 2016;
- “Advances in Formal Models of Search and Search Behavior” Tutorial, Prof. Leif Azzopardi from University of Glasgow and Prof. Guido Zuccon from Queensland University of Technology, Melbourne, December 5, 2016;

- WSDM 2017 Doctoral Consortium, Eugene Agichtein from Emory University and Yoelle Maarek from Yahoo Research, Cambridge, February 6, 2017;
- “Neural Networks for Information Retrieval (NN4IR)” Tutorial, Tom Kenter, Alexey Borisov, Christophe Van Gysel, Mostafa Dehghani, Maarten de Rijke from University of Amsterdam, and Bhaskar Mitra from Microsoft, University College London, Tokyo, August 7, 2017.

Seminars

- “Future Internet Security and Privacy (challenges)”, Prof. Mauro Conti e Dr. Moreno Ambrosin from Università degli studi di Padova, October 31, 2015.
- “Hybrid Approaches for Synchrony and Memory for Parallel Graph Algorithms”, Prof. Nancy Amato from Texas A&M University, July 17, 2015.
- “Evaluation Effort, Reliability and Reusability in XML Retrieval”, Prof. Sukomal Pal, from Indian School of Mines, October 21, 2015;
- “Arithmetic for Rooted Trees”, Prof. Fabrizio Luccio, Professore Emerito at Università di Pisa, November 6, 2015;
- “Vito Volterra e lo studio della complessità delle interazioni ecologiche”, Prof. Marino Gatto from Politecnico di Milano, May 17, 2016;
- “Single and Multiobjective clustering”, Prof. Ujjwal Maulik from the Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, May 18, 2016;
- “When is Big Data Sufficiently Big? When is it Too Big? Sample Complexity, Uniform Convergence, and Generalization Error”, Prof. Eli Upfal from Brown University, Providence, USA, June 6, 2016;
- “Large Scale Data Analytics: Challenges, and the role of Stratified Data Placement”, Prof. Srinivasan Parthasarathy from Ohio State University, USA, 16 June 2016;
- “Google Maps: an informal discussion”, Hugh E. Williams, vice president at Google, August 4, 2016;
- “International Students Information Seeking Behaviour: Implications for International Education”, Dr. Catherine Gomes, senior lecturer at RMIT University, August 11, 2016;
- Melbourne Search August 2016 Meetup: “Evaluating the Evaluation: Measuring things that Matter”, Prof. Alistair Moffat from University of Melbourne; “Counterfactual Evaluation: Estimating New User Behaviour from Historical Data”, Prof. Mark Sanderson from RMIT University; “Impact of search effectiveness on SEO: A case study in Marketplace”, Anuj Luthra and Sadegh Kharazmi from Redbubble, August 31, 2016;
- “An Overview of Scientific Text Mining and Summarization Research at TALN”, Prof. Horacio Saggion from the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, September 8, 2016;

- “KRNN: k rare-class nearest neighbour classification” and “Automatic labelling of topics via analysis of user summaries”, Prof. Xiuzhen Zhang from RMIT University, September 15, 2016;
- “Neural Net Approaches to Information Retrieval”, Prof. Bruce Croft, from University of Massachusetts Amherst, October 6, 2016;
- “Pricing Relational Data with Formal Guarantees”, Prof. Paris Koutris, University of Wisconsin-Madison, March 24, 2017;
- “Is Computer Science Dying?”, Prof. Alex Nicolau, from University of California, Irvine, July 20, 2017;
- “Text, Computer, and Style: In the Pursuit of Elena”, Prof. Jacques Savoy from University of Neuchatel, September 6, 2017;
- “Personal Behavioral Technology - Wearables Can Become an Active Contributor to Your Wellbeing”, Prof. Chris Van Hoof, from Holts Centre/IMEC, September 22, 2017.

Participation to National and International Conferences and Workshops

- The 5th Conference and Labs of the Evaluation Forum, Information Access Evaluation meets Multilinguality, Multimodality, and Interaction, (CLEF 2014), Sheffield (UK), September 15-18, 2014;
- The 6th Italian Information Retrieval Workshop (IIR2015), Cagliari, May 25-26, 2015;
- Postdoctoral Research in Informatics, at the Department of Information Engineering of the University of Padua, Padova, July 8, 2015;
- The 6th BCS-IRSG Symposium on Future Directions in Information Access, (FDIA 2015), Thessaloniki, Greece, September 2, 2015;
- The 1st ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR2015), Northampton Massachusetts (USA), September 27-30, 2015;
- The 7th Italian Information Retrieval Workshop, Venezia, Italy, May 30- 31, 2016;
- The 21st Australasian Document Computing Symposium (ADCS 2016), Melbourne, Australia, December 6-7, 2016;
- The 10th ACM International Conference on Web Search and Data Mining (WSDM 2017), Cambridge, UK, February 6-10, 2017;
- “WSDM 2017 Workshop on Mining Online Health Reports”, Nigel Collier and Nut Limsopatham from University of Cambridge, Ingemar J. Cox and Vasileios Lampos from University College London, Aron Culotta from Illinois Institute of Technology, and Mike Conway from University of Utah, Cambridge, February 10, 2017;
- The 8th Italian Information Retrieval Workshop (IIR 2017), Lugano, Switzerland, June 5-7, 2017;

- The 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, (SIGIR 2017), Tokyo, Japan, August 7-11, 2017;
- “ECOM’17: eCommerce” Workshop, Jon Degenhardt from eBay, Surya Kallumadi from Kansas State University, Maarten de Rijke from University of Amsterdam, Luo Si from Alibaba, Andrew Trotman from University of Otago, and Xu Yinghui from Taobao, Tokyo, August 11, 2017.

Other Activities

- **Teaching**
 - Teaching of Mathematics with element of Statistics, Bachelor Course in Natural Science, Department of Biology, University of Padua, October 2014, 30h;
 - Tutor of Mathematics with element of Statistics, Bachelor Course in Natural Science, Department of Biology, University of Padua, from October 2014 to February 2015, 25h;
 - Tutor of Mathematics with element of Statistics, Bachelor Course in Natural Science, Department of Biology, University of Padua, from October 2015 to February 2016, 25h;
 - Tutor of Database Management Systems, Master degree in Computer Engineering, Department of Information Engineering, University of Padua, from March 2016 to May 2016, 14h;
 - Tutor of Database Management Systems, Master degree in Computer Engineering, Department of Information Engineering, University of Padua, from March 2017 to May 2017, 20h.
- **Professional Activity as a Reviewer**
 - **International Journals**
 - Information Processing and Management (IP&M)
 - **International Conferences**
 - Short paper PC member of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016). Pisa, Italy, July 17-21, 2016;
 - Short paper PC member of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017). Tokyo, Japan, August 7-11, 2017;
 - Short paper PC member of the 26th ACM International Conference on Information and Knowledge Management (CIKM 2017). Singapore, November 6-10, 2017.
 - **International Workshops**
 - Full paper PC member of the 8th International Workshop on Evaluating Information Access (EVIA 2017). Tokyo, Japan, December 5, 2017.

- **Organization of Scientific Events**

- 8th European Conference on Information Retrieval (ECIR 2016). Padova, Italy, March 20-23, 2016, Role: Member of the Organization Committee;
- 1st International Workshop on LEARning Next gEneration Rankers (LEARNER 2017). Amsterdam, Netherlands, October 1, 2017, Role: Organizer.

- **Award**

- Best paper award for the paper: Di Nunzio, G. M., Maistro, M., and Zilio, D. (2016). Gamification for machine learning: The classification game. In Hopfgartner, F., Kazai, G., Kruschwitz, U., and Meder, M., editors, *Proceedings of the Third International Workshop on Gamification for Information Retrieval co-located with 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016), Pisa, Italy, July 21, 2016*, volume 1642 of CEUR Workshop Proceedings, pages 45–52. CEUR-WS.org.

Mobility periods

- June 12– 17, 2016: visiting student at Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, (ISTI), CNR, Pisa, collaborating with Dr. Claudio Lucchese and Dr. Raffaele Perego on possible ways to extend learning to rank algorithms by integrating the user dynamic;
- July 2016 – January 2017: visiting student at the School of Computer Science and Information Technology at RMIT, Melbourne, collaborating with Prof. Mark Sanderson on analyzing job seekers behavior and possible approaches to adapt click models to job search;
- July 2016 – January 2017: internship at SEEK, company which developed and maintains a job search engine in Melbourne, collaborating with Dr. Wilson Wong and Dr. Sargol Sadeghi on analyzing job seekers behavior and possible approaches to adapt click models to job search;
- February 2016 – Now: scientific collaboration with Dr. Raffaele Perego and Dr. Claudio Lucchese from Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, (ISTI), CNR, Pisa, collaborating with Dr. Claudio Lucchese and Dr. Raffaele Perego on possible ways to extend learning to rank algorithms by integrating the user dynamic;

PART 2 - RESEARCH ACTIVITY

My Ph.D. topic deals with Information Retrieval (IR), i.e. the activity of retrieving and presenting relevant information from a collection of documents in order to satisfy the user’s information need. In particular, IR is concerned with complex systems delivering a variety of key applications to industry and society: Web search engines, (bio)medical search, enterprise search, intellectual property and patent search, expertise retrieval systems, and many others.

My research activity in IR can be summarized in the following main points:

1) *Formal Framework for Information Retrieval Evaluation Measures:*

We presented a formal framework to define and study the properties of utility-oriented measurements of retrieval effectiveness. The proposed framework is laid in the wake of the representational theory of measurement, which provides the foundations of the modern theory of measurement in both physical and social sciences, thus contributing to explicitly link IR evaluation to a broader context. The proposed framework contributes to a better understanding and a clear separation of what issues are due to the inherent problems in comparing systems in terms of retrieval effectiveness and what others are due to the expected numerical properties of a measurement.

2) *Markovian User Models:*

We proposed a family of new evaluation measures, called Markov Precision (MP), which exploits continuous-time and discrete-time Markov chains in order to define a new user model. Continuous-time MP behaves like time-calibrated measures, bringing the time spent by the user into the evaluation of a system; discrete-time MP behaves like traditional evaluation measures. We also proposed several alternative models that take into account different possible behaviors in scanning a ranked result list.

Moreover, we will explore how this model based on Markov chains may represent a way to exploit user personalization. For example, we will classify users on the basis of their movements on the result list of documents and infer their typical behavior, e.g. navigational or informative users. We will also estimate relevance of documents by analyzing invariant distributions shapes, i.e. a peak in the probability distribution may represent a cluster of relevant documents in the corresponding run.

3) *Integration of Markovian User Models in Learning to Rank:*

We aim at improving Learning to Rank algorithms by letting them to take into account user interactions to increase their overall effectiveness. In particular, we developed a model of user behavior where the interaction between the user and the ranked result list is modeled via Markov chains. Then, we integrated the proposed model into LambdaMART, a Learning to Rank algorithm, in order to take into account the user interactions when the algorithm learns the scoring function.

We evaluated the performance of the proposed approach on publicly available LtR datasets, finding that the improvements measured over the standard algorithm are statistically significant.

Finally, we will organize the workshop LEARNER, co-located with ICTIR2017. The goal of this workshop is to investigate how to improve ranking, in particular LtR, by bringing in new perspectives which have not explored or fully addressed yet, for example embedding user behaviour and dynamic in LtR, incremental LtR, online, or personalized LtR, unsupervised approaches to LtR, active learning for LtR, transfer learning for LtR and many others.

4) *Identifying the Differences between Job and Web Search Users:*

We analyzed clicks and query logs for job and talent search from a popular employment marketplace. The observations suggest that the understanding of users' search behavior in this scenario is still at its infancy and that some of the assumptions made in general web search may not hold true. We compared users' search behavior for job search, talent search and more traditional

web-search. Our purpose is to better understand whether the underlying assumptions we have with regard to user models, ranking factors and success metrics in web search can (or should) hold true for job and talent search. Moreover, our aim is to understand whether a Markovian user model can be learnt and fitted to job search users in order to describe their behavior.

5) *Evaluation of Click Models:*

We find that the click prediction evaluation measures proposed in literature do not evaluate the models correctly, in particular they are not sensitive to the imbalance between the number of clicks and skips. We propose a formal explanation to clarify why one measure, perplexity, does not take into account the imbalance between click and skips. We highlighted this issue with two different use cases: Web search, when considering low rank positions documents receive just a few clicks, and job search, users tend to click less in job search compared to web search. We will test different measures based on click and skip perplexities, classification measures and other probabilistic approaches to address this issue. Finally, we will compare the proposed measures against state of the art measures, on Web and job search datasets, and we will study their formal properties and robustness to clicks and skips unbalance.

6) *Gamification for Machine Learning:*

In recent works, it has been shown that a mix of crowdsourcing and active learning approaches can be used to annotate objects at an affordable cost. We explored the gamification of machine learning techniques; in particular, the problem of classification of objects. We conducted a first pilot study by designing a simple game, based on a visual interpretation of probabilistic classifiers, that consists in separating two sets of coloured points on a two-dimensional plane by means of a straight line. We compared the accuracy of the player to the state-of-the-art classification algorithms, showing that potentially, the gamification problem may give a strong indication about how to stop the labelling process and use the annotated dataset to train with a very high accuracy state-of-the-art algorithm. Moreover, we used this gamified classifier to that mimic the behaviour of a user that tries to find the optimal decision with two different strategies and participate in the Total Recall Track of TREC 2016.

7) *Approach to Merge Relevance Judgements from Crowd Source Assessors:*

We proposed the Assessor-driven Weighted Averages for Retrieval Evaluation (AWARE) framework, a novel methodology for dealing with multiple assessments which may be contradictory and/or noisy. By modelling assessors as a random variable, AWARE takes the expectation of a generic evaluation measure composed with this random variable. In this way, it approaches the problem of aggregating different assessors from a new perspective, i.e. directly combining the performance scores computed on the ground-truth generated by the different assessors instead of adopting some machine learning approach to classify the labels produced by the assessors themselves. We proposed new ways of weighting assessors based on the distance between them and random assessors and we compared our approach with existing approaches.

PART 3 - PUBLICATIONS

List of publications on international journals

- J1. Ferrante, M., Ferro, N., and Maistro, M. (2017). AWARE: Exploiting Evaluation Measures to Combine Multiple Assessors. *ACM Transactions on Information Systems (TOIS)*, 36, 2, Article 20, 38 pages.

List of publications on conference proceedings

- C1. Ferrante, M., Ferro, N., and Maistro, M. (2014). Rethinking how to Extend Average Precision to Graded Relevance. In Kanoulas, E., Lupu, M., Clough, P. D., Sanderson, M., Hall, M. M., Hanbury, A., and Toms, E. G., editors, *Information Access Evaluation. Multilinguality, Multimodality, and Interaction - 5th International Conference of the CLEF Initiative, (CLEF 2014)*, Sheffield, UK, September 15-18, 2014. *Proceedings*, volume 8685 of *Lecture Notes in Computer Science*, pages 19–30. Springer.
- C2. Ferrante, M., Ferro, N., and Maistro, M. (2014). Injecting User Models and Time into Precision via Markov Chains. In Geva, S., Trotman, A., Bruza, P., Clarke, C. L. A., and Järvelin, K., editors, *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, (SIGIR 2014)*, Gold Coast, QLD, Australia, July 6-11, 2014, pages 597–606. ACM.
- C3. Ferrante, M., Ferro, N., and Maistro, M. (2015). Towards a Formal Framework for Utility-Oriented Measurements of Retrieval Effectiveness. In Allan, J., Croft, W. B., de Vries, A. P., and Zhai, C., editors, *Proceedings of the 1st International Conference on The Theory of Information Retrieval, (ICTIR 2015)*, Northampton, Massachusetts, USA, September 27-30, 2015, pages 21–30. ACM.
- C4. Di Nunzio, G. M., Maistro, M., and Zilio, D. (2016). The University of Padua (IMS) at TREC 2016 Total Recall Track. *Proceedings of the 25th Text REtrieval Conference (TREC 2016)*, Gaithersburg, Maryland, USA, November 15-18, 2016.
- C5. Maistro, M. (2017). Adapting Information Retrieval to User Signals via Stochastic Models. In de Rijke, M. and Shokouhi, M. and Tomkins, A., and Zhang, M., editors, *Proceedings of the 10th ACM International Conference on Web Search and Data Mining, (WSDM 2017)*, Cambridge, United Kingdom, February 6-11 2017, pages 843–843, ACM.
- C6. Ferro, N., Lucchese, C., Maistro, M., and Perego, R. (2017). On Including the User Dynamic in Learning to Rank. In Kando, N., Sakai, T., Joho, H., Li, H., de Vries, A. P. and White, R., editors, *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, (SIGIR 2017)*, Tokyo, Japan, August 7-11, 2017, ACM.
- C7. Ferro, N., Lucchese, C., Maistro, M., and Perego, R. (Accepted in 2017). LEARning Next gEneration Rankers (LEARNER 2017). In Kamps, J., Kanoulas, E., de Rijke, M., Yilmaz, E., and Fang, H., editors, *Proceedings of the 3rd International Conference on The Theory of Information Retrieval, (ICTIR 2017)*, Amsterdam, The Netherlands, October 1-4, 2017. ACM.

List of publications on workshop proceedings

- W1. Ferrante, M., Ferro, N., and Maistro, M. (2015). Markov precision: Modelling user Behaviour over Rank and Time. In Boldi, P., Perego, R., and Sebastiani, F., editors, Proceedings of the 6th Italian Information Retrieval Workshop, (IIR 2015), Cagliari, Italy, May 25-26, 2015, volume 1404 of CEUR Workshop Proceedings. CEUR-WS.org.
- W2. Maistro, M. (2015). Improving Information Retrieval Evaluation via Markovian User Models and Visual Analytics. In Azzopardi, L. and Wilson, M., editors, Proceedings of the 6th BCS-IRSG Symposium on Future Directions in Information Access, (FDIA 2015), Thessaloniki, Greece, August 31 - September 4, 2015, Workshops in Computing. BCS.
- W3. Di Buccio, E., Di Nunzio, G. M., Ferro, N., Harman, D., Maistro, M., and Silvello G. (2015). Unfolding Off-the-shelf IR Systems for Reproducibility. In Arguello, J., Diaz, F., Lin, J., and Trotman, A., editors, Proceedings of the First Workshop on Reproducibility, Inexplicability, and Generalizability of Results (RIGOR) co-located with 38th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (SIGIR 2015), Santiago, Chile.
- W4. Ferro, N., Ferrante, M., and Maistro, M. (2016). Basis of a Formal Framework for Information Retrieval Evaluation Measurements. In Di Nunzio, G. M., Nardini, F. M., and Orlando, S., editors, Proceedings of the 7th Italian Information Retrieval Workshop, (IIR 2016), Venezia, Italy, May 30-31, 2016, volume 1653 of CEUR Workshop Proceedings. CEUR-WS.org.
- W5. Di Nunzio, G. M., Maistro, M., and Zilio, D. (2016). Gamification for Machine Learning: The Classification Game. In Hopfgartner, F., Kazai, G., Kruschwitz, U., and Meder, M., editors, Proceedings of the Third International Workshop on Gamification for Information Retrieval co-located with 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016), Pisa, Italy, July 21, 2016, volume 1642 of CEUR Workshop Proceedings, pages 45–52. CEUR-WS.org.
- W6. Di Nunzio, G. M., Maistro, M., and Zilio, D. (2016). Gamification for Information Retrieval: The Query Aspects Game. In Basile, P., and Corazza, A., and Cutugno, F., and Montemagni, S., and Nissim, M., and Patti, V. and Semeraro, G., and Sprugnoli, R., editors, Proceeding of the 3rd Italian Conference on Computational Linguistics, (CLiC-it 2016), Napoli, Italy, December 5-6, 2016, volume 1749 of CEUR Workshop Proceedings. CEUR-WS.org.
- W7. Di Nunzio, G. M., Maistro, M., and Zilio, D. (2017). A Game of Lines: Developing Game Mechanics for Text Classification. In Crestani, F., Di Noia, T., and Perego, R., editors, Proceedings of the 8th Italian Information Retrieval Workshop, (IIR 2017), Lugano, Switzerland, June 5-7, 2017, CEUR Workshop Proceedings. CEUR-WS.org.
- W8. Spina, D., Maistro, M., Ren, Y., Sadeghi, S., Wong, W., Baldwin, T., Cavedon, L., Moffat, A., Sanderson, M., Scholer, F., and Zobel, J. (2017). Understanding User Behavior in Job and Talent Search: An Initial Investigation. In Degenhardt, J., Kallumadi, S., de Rijke, M., Si, L., Trotman, A., and Yinghui X., editors, Proceedings of the 2017 SIGIR workshop on eCommerce, (eCom 2017), co-located with the 40th International ACM SIGIR Conference on

Research and Development in Information Retrieval (SIGIR 2017), Tokyo, Japan, August 11, 2017, CEUR Workshop Proceedings. CEUR-WS.org.

September 26, 2017

Student signature

Maria Chiostro

Supervisor signature

W. Chiostro