

Curriculum Vitae

1 Personal Information

Name Ricardo Baeza-Yates
E-mail rbaeza@acm.org
Personal website <http://www.baeza.cl/>

2 Current Activities

- Research Professor (half-time) at the Institute for Experiential AI, Northeastern University, since 2021.
- Consultant for startups, companies, and venture capital in Brazil, Chile, Mexico, Spain, and USA; as well as for international organizations like United Nations and IADB (BID).
- Member of Global Partnership of AI (Data Governance Working Group), Spain's Advisory Council of AI, Catalonia's AI Observatory Advisory Board, ACM's Policy Technology subcommittee in AI and Algorithms, and IABD's Advisory Committee for fAIr LAC (Fair AI in Latin America and the Caribbean).
- Full Professor (part-time) at the Dept. of Information and Communication Technologies of Universitat Pompeu Fabra, Barcelona, Catalonia, Spain, since 2006.
- Full Professor (part-time) at the Dept. of Computing Sciences, Universidad de Chile, Santiago, Chile, since 1989.
- Corresponding member of the Brazilian Academy of Sciences since 2019.
- Founding member of the Chilean Academy of Engineering since 2010.
- Corresponding member of the Chilean Academy of Sciences since December 2002.
- Member of the advisory board of the National Portuguese Research Lab in CS, INESC, since 2012.
- Member of the advisory board of the CSE Department of UC Riverside as well as the Department of Informatics Engineering, University of Porto, Portugal, since 2020.
- Member of the Advisory Board of the Professional Association of Computing Professionals of Catalonia since 2009.

3 Highlights

- World class expert in search, data mining and data science.
- Research innovator and team builder with leadership and multitasking skills.

- VP of Research for Yahoo Labs (2006-2016) with experience as research manager and technical lead in applied projects on web search, usage data mining and web advertising (sponsored search and native).
- Co-author of the most used and cited textbook on search concepts and technology.
- ACM (less than 1% of the members) and IEEE Fellow.

4 Studies

- Ph.D. in Computer Science, University of Waterloo, Ontario, Canada, May 1989. Supervised by Gaston Gonnet, and working in the New Oxford English Dictionary Project.
- M.Eng. in Electrical Engineering (Digital Systems and Automatic Control), Univ. de Chile, 1986.
- M.Sc. in Computer Science, Universidad de Chile, 1985.
- Electrical Engineer, specialization in Digital Systems, Universidad de Chile, 1985.
- Bachelor in Computer Science, Universidad de Chile, 1983.

5 Awards, Distinctions, Fellowships and Scholarships

- Named Corresponding Member of the Brazilian Academy of Sciences, 2019.
- Award Salvà i Campillo to the Person of the Year, Catalan Association of Telecommunications Engineers with the cooperation of the Association of Computing Engineers of Catalonia, 2019.
- National Research Award for applied research and technology transfer, Scientific Computing Societies of Spain & BBVA Foundation, Spain, 2018.
- Best paper award at Techpulse 2014, the main annual technical event of Yahoo. This award was obtained with Di Jiang, Fabrizio Silvestri and Beverly Harrison for a new algorithm that using deep learning allows to predict the next mobile app that a person will use with more than 90% accuracy.
- Computer World Spain Award to Yahoo Labs Barcelona for contributions to big data in 2014.
- Master inventor award, Yahoo, 2012.
- 2012 ASIS&T Best Information Science Book Award for the second edition of Modern Information Retrieval (see books).
- Named IEEE Fellow for contributions to the development of computer science, 2011.
- National Award of the Chilean Association of Engineers, 2010.
- Founding member of the Chilean Academy of Engineering, 2010.
- Named ACM Fellow for contributions to information retrieval algorithms and techniques, 2009.

- CLEI Distinction to Achievements in Informatics, given by the Latin American association of CS departments to people that has contributed to the development of CS in the region, 2009.
- J. W. Graham Medal in Computing & Innovation given to distinguished alumni of the University of Waterloo, Canada, 2007.
- Erskine Fellowship from the University of Canterbury, New Zealand, 2004.
- Chilean Fulbright Commission Fellowship in Information Technology, 2003.
- Manuel Montt award to “Modern Information Retrieval” (co-authored with Berthier Ribeiro-Neto) as the best technical book published by Chilean authors during 1997-2001 (shared with Tomás Moulian’s essay “Current Chile: Anatomy of a Myth”), 2002.
- Named corresponding member of the Chilean Academy of Sciences, 2002.
- Recognition from Informatics Magazine (IDG Chile) for the Development of TIs in Chile, 2001.
- COMPAQ Award to the best Brazilian research article in CS published in 1996 (with Eduardo Barbosa and Nivio Ziviani), 1997.
- Fellowship from the Education and Science Ministry of Spain to have a sabbatical year at the Polytechnic Univ. of Catalunya, 1996.
- Award “Ramón Salas Edwards” of the Institute of Engineers of Chile to the best research on engineering on the last 4 years in 1994.
- Recipient of a United States Information Service Important Visitor program, June 1994.
- Award of the Organization of American States (OAS) for young researchers in 1993 (Exact Sciences).
- Award from PC Software magazine to the best Chilean software, SearchCity, designed together with Pablo Palma in 1992. This software for Windows 3.1 was similar to Google Desktop but was too ahead in time to succeed.
- Award “Marcos Orrego Puelma” of the Institute of Engineers of Chile to the best engineering student graduated in Chile during 1985.
- Five PhD scholarships from Ontario, Univ. of Waterloo, ICR and ITRC, May 1986-April 1989.

6 Professional and Entrepreneurship Experience

- CEO of NTENT Hispania, a R&D subsidiary of NTENT USA, Barcelona, Spain (April 2017 - December 2020).
- CTO of NTENT, a semantic search technology company, Carlsbad, USA (June 2016 - June 2020).
- Chief Research Scientist of Yahoo Labs from October 2015 to February 2016 in Sunnyvale, USA.

- VP of Research at Yahoo Inc. from August 2014 to February 2016 in Sunnyvale, USA.
- VP of Research for EMEA and Latin America at Yahoo Inc. from January 2006 to September 2015, where I started and lead the labs in Europe (Barcelona, Spain) and Latin America (Santiago, Chile).
- Member of the Scientific Advisory board of EURECAT, Barcelona (2015-2017).
- Member of the Executive Board of Barcelona Media Foundation (2010-2015).
- Member of the board of directors of Yahoo Hungary (2009-2016) and Yahoo Israel (2009-2012).
- Supervised Yahoo Labs London from its inception in 2012 and until 2015.
- Supervised Yahoo Labs Haifa from its inception in 2008 and until 2012.
- Co-founder of Raditech in 2001, a Spanish company that offered vertical search as a service, particularly Buscopio, a Spanish web search engine. The company was sold to Fonetic, another Spanish company, in 2007.
- Founder and CEO of TodoCL, a web search engine that covered Chilean websites from 2000 to 2015. It had more than 100 thousand indexed websites and millions of queries per month.
- Member of the consulting committee for the design and evaluation of the new identity card and passport for the National Chilean Civil Registry from 1995 to 2005. In particular I was in charge of evaluating the AFIS (Automated Fingerprint Identification System) system for identity verification.
- International consultant for the United Nations (PNUD) for a governmental system in Perú, 1998.
- Consultant for the International Organization for Migrations, United Nations, for the analysis of the information systems of the Chilean Congress, 1990.
- Consultant for many companies in Chile, including IBM, Langton-Clarke, Sonda, etc., 1990-1992.

7 Academic Experience

- Director for Graduate Programs in Data Science (half-time), Northeastern University at Silicon Valley (2018-2020).
- ICREA Research Professor, Dept. of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain, Dec 2004-Dec 2005. Part-time Professor since 2006.
- Faculty member at the Department of Computer Science, University of Chile, Santiago, Chile, 1985-2005 (Professor since April 1995). Part-time Professor since 2006.
- Adjunct Professor, Dept. of Computer Science, University of Waterloo, Canada, June 2016-today.
- Founder and first director of the Center for Web Research (www.cwr.cl), 2002-2005.
- Chair of the CS Dept. (Univ. of Chile) from June 1993 to September 1995, and also from November 2002 to November 2004.

8 Patents

1. Ricardo Baeza-Yates. Techniques for Searching Future Events, Yahoo! Inc. and Univ. of Chile, August 2006. USA Patent #7,668,813, granted Feb 23, 2010.
2. Ricardo Baeza-Yates and Barbara Poblete. User Query Data Mining and Related Techniques, Yahoo! Inc. and Univ. of Chile, September 2006. USA Patent #7,617,208, granted Nov 10, 2009.
3. Ricardo Baeza-Yates, Aristides Gionis, Flavio Junqueira and Vassilis Plachouras. System and Method for Caching Posting Lists, Yahoo! Inc., October 2007. USA Patent #7,890,488, granted Feb 5, 2011.
4. Ricardo Baeza-Yates, Alessandro Tiberi. Extracting Semantic Relations from Query Logs, Yahoo! Inc., December 2007. USA Patent #7,895,235, granted Feb 22, 2011.
5. Luciano Barbosa, Flavio Junqueira, Vassilis Plachouras, Ricardo Baeza-Yates. Method and System for Quantifying the Quality of Search Results based on Cohesion, Yahoo! Inc., December 2007. USA patent #7,720,870, granted May 18, 2010.
6. Barbara Poblete, Ricardo Baeza-Yates. Classifying Documents using Implicit Feedback and Query Patterns, Yahoo! Inc., July 2008. USA patent # 8,645,369, granted February 4, 2014.
7. Ricardo Baeza-Yates, Aristidis Gionis, Flavio Junqueira, Vassilis Plachouras and Luca Tello. System and Methodology for a Multi-Site Search Engine, Yahoo! Inc., October 2008. USA patent #8,095,545, granted January 10, 2012.
8. Barbara Poblete, Ricardo Baeza-Yates. Methods for Web Site Analysis, Yahoo! Inc., April 2010. USA patent #8,751,632, granted June 10, 2014.
9. Fabrizio Silvestri, Ricardo Baeza-Yates, Beverly Harrison, Di Jiang. Predicting the Next Application that You are going to Use on Aviate, Yahoo! Inc., December 2014. USA Patent 9971972B2, granted May 15, 2018.
10. Manjana Chandrasekharan, Keiko Horiguchi, Amanda Stent, Ricardo Baeza-Yates, Jeffrey Kuwano, Achint Thomas, Yi Chang. Audio Verification, Yahoo! Inc., September 2015. USA Patent #10277581B2, granted April 30, 2019.
11. Lizi Zhang, Hongbo Deng, Amit Goyal, Yi Chang, Ricardo Baeza-Yates. Computerized System and Method for Search Query Auto-completion, Yahoo! Inc, September 2015. USA Patent #10049149B2, granted August 14, 2018.
12. Simon Osindero, Ioannis Kalantidis, David Shamma, Lyndon Kennedy, Ayman Farahat, Gerry Pesavento, Ricardo Baeza-Yates. Method and System for Selecting Supplemental Content using Visual Appearance, Yahoo! Inc, April 2016.

9 Research Projects

At Yahoo Labs I was involved in many internal applied research projects in several areas, including web search, computational advertising (sponsored search and native), web usage data mining, and social media analytics. They included many aspects, in particular efficiency and scalability.

I have coordinated or participated in EU Research Projects, Ibero-American Research Projects and National Spanish and Chilean Research Projects.

10 Citations

According to Google Scholar there are more than 48,500 citations to more than 600 publications found (December 2020), with an h-index of 85. Citations below are mainly from this source (see my personal entry). I am the most cited in algorithms and data structures, 2nd most cited in web search, the 3rd most cited in web mining, and 11th most cited in information retrieval.

In December 2020, the ACM Digital Library had more than 7,700 citations to more than 310 publications, with more than 11,700 downloads in the previous 12 months and 93,500 overall.

In May 2012, CiteSeerX found more than 5,380 independent citations and ranked me the 528th most cited author in CS from a universe of more than 800 thousand authors (since then has not been updated). In their statistics my information retrieval book is the third most cited work of 1999 and the 86th most cited overall (2015).

On March 2020, Microsoft Academic Search had 487 of my publications with almost 44,000 citations. I appear in 16th place in information retrieval and 42nd place in web related research.

11 Selected Publications

Only recent, applied or well cited works are considered among more than 600 publications.

11.1 Books

- [1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval: the concepts and technology behind search*, second edition. Addison-Wesley, England, 913 pages, 2011. This edition won the 2012 ASIS&T Book of the Year award and has been already translated to Chinese (2012, 672 pages) and the 10 main chapters to Portuguese (2013, 614 pages). Part of the book is available at mir2ed.org. The first edition was published by ACM Press/Addison-Wesley, England, 513 pages, in 1999, and was translated to Chinese and Korean and printed in special editions for China and India. According to CiteseerX this is the 3th most cited CS work published in 1999 and the 67th overall. More than **13,800** citations.
- [2] B.B. Cambazoglu and R. Baeza-Yates. *Scalability Challenges in Web Search Engines*, in Synthesis Lectures on Information Concepts, Retrieval, and Services (Gary Marchionini, editor). Morgan & Claypool Publishers, USA, 122 pages, 2015.
- [3] G.H. Gonnet and R. Baeza-Yates. *Handbook of Algorithms and Data Structures - In Pascal and C*. Addison-Wesley, Wokingham, UK, 424 pages, 1991. (second edition). Source code is available in my personal WWW page. More than **780** citations.

- [4] W. Frakes and R. Baeza-Yates, editors. *Information Retrieval: Data Structures and Algorithms*. Prentice-Hall, Englewood Cliffs, NJ, USA, 504 pages, 1992. Source code is available in my personal WWW page. More than **2,500** citations.
- [5] M. Melucci and R. Baeza-Yates, editors. *Advanced Information Retrieval*, Springer, Berlin, 2011. Based in the European IR summer school organized in Padova, 2009.
- [6] I. King and R. Baeza-Yates, editors. *Weaving Services and People on the World Wide Web*, Springer, Berlin, 2009. Selection of papers presented at the Workshop Track of the 17th International World Wide Web Conference (WWW 2008) held at Beijing, China, 2008.
- [7] R. Baeza-Yates, J. Glaz, H. Gzyl, J. Hüsler, and J.L. Palacios, editors. *Applied Probability: Recent Advances*, Kluwer Academic Publishers, 2004.

11.2 Chapters in Books

- [1] R. Baeza-Yates, R. Blanco, M. Castellanos. Web Text Mining. In *The Oxford Handbook of Computational Linguistics*, 2nd edition, R. Mitkow, editor. Oxford Press, 2017.
- [2] J. Lehmann, M. Lalmas, and R. Baeza-Yates. Measuring Inter-Site Engagement. In *Handbook of Statistics: Big Data Analytics*, V. Govindaraju, V. V. Raghavan and C. R. Rao, editors. Volume 33. Elsevier, 2015.
- [3] R. Baeza-Yates and Y. Maarek. Usage Data in Web Search: Benefits and Limitations. In *Scientific and Statistical Database Management: 24th SSDBM*, A. Ailamaki and S. Bowers (eds). LNCS 7338, Springer, Chania, Crete, 495–506, June 2012.
- [4] R. Baeza-Yates, A. Broder and Y. Maarek. The new frontier of Web Search Technology: Seven challenges. In *Search Computing – Trends and Developments*, S. Ceri and M. Brambilla (eds). LNCS 6585, Springer, Berlin, 2011.
- [5] Ricardo Baeza-Yates, Paolo Boldi, Alessandro Bozzon, Marco Brambilla, Stefano Ceri, Gabriella Pasi. Trends in Search Interaction. In *Search Computing – Trends and Developments*, S. Ceri and M. Brambilla (eds). LNCS 6585, Springer, Berlin, 2011.
- [6] Ricardo Baeza-Yates, Paolo Boldi. Web Structure Mining. In *Advanced Techniques in Web Intelligence*, edited by Juan Velásquez and Lakhmi Jain, Springer, pp. 113-142, 2010.
- [7] Ricardo Baeza-Yates. Privacy in Query Logs. In *Web Intelligence and Security*, edited by M. Last and A. Kandel. IOS Press, pp. 189-193, 2010.
- [8] Ricardo Baeza-Yates, Rosie Jones, Barbara Pobleto, Myra Spiliopoulou. Issues with Privacy Preservation in Query Log Mining. In *Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques*, edited by Francesco Bonchi and Elena Ferrari. Chapman and Hall/CRC Press, pp. 353–368, 2010.
- [9] Ricardo Baeza-Yates, Prabhakar Raghavan. Next Generation Web Search. In *Search Computing - Challenges and Directions*, edited by S. Ceri and M. Brambilla, LNCS 5950, Springer, pp. 11-23, 2010.

- [10] Ricardo Baeza-Yates and Barbara Poblete. A Website Mining Model Centered on User Queries. In *Semantics, Web and Mining*, M. Ackermann et al. (Eds.): EWMF/KDO 2005, Springer LNAI 4289, pp. 1–17, 2006.
- [11] Ricardo Baeza-Yates and Carlos Castillo. Web Searching. In *Encyclopedia of Language and Linguistics*, 2nd edition. Graeme Hirst (editor), Vol. 13, Elsevier: Oxford, 2005, 527-537.
- [12] R. Baeza-Yates and G. Navarro. Text Searching: Theory and Practice. In *Formal Languages and Applications*, Carlos Martin-Vide, Victor Mitran, Gheorghe Paun (Eds.), Studies in Fuzziness and Soft Computing 148, Springer, Berlin, 2004, 565–597.
- [13] Ricardo Baeza-Yates. Query Usage Mining in Search Engines. In *Web Mining: Applications and Techniques*, Anthony Scime, editor. Idea Group, 2004, 307–321. More than **70** citations.
- [14] Ricardo Baeza-Yates, Carlos Castillo and Felipe Saint-Jean. Web Dynamics, Structure and Page Quality. In Mark Levene and Alex Poulouvasilis (editors), *Web Dynamics*, Springer Verlag, 2004, 93–109.
- [15] Ricardo Baeza-Yates, Benjamin Bustos, Edgar Chavez, Norma Herrera, and Gonzalo Navarro. Clustering in Metric Spaces and Its Application to Information Retrieval. In Weili Wu and Hui Xiong (editors), *Clustering and Information Retrieval*. Kluwer Academic Publishers, 2003, 1–34.
- [16] Ricardo Baeza-Yates, Alistair Moffat and Gonzalo Navarro. Searching Large Text Collections. In J. Abello, P. Pardalos and M. Resende (editors), *Handbook of Massive Data Sets*, Kluwer Academic Publishers, 2002, pages 195-244.
- [17] R. Baeza-Yates and P. Poblete. Searching. In Mikhail Atallah, editor, *Handbook on Algorithms and Theory of Computation*, chapter 2. CRC Press, 1999. A revised version appears in the first volume of the second edition published in 2009 and also edited by Marina Blanton.
- [18] R. Baeza-Yates. Searching: An algorithmic tour. In Allen Kent and James G. Williams, editors, *Encyclopedia of Computer Science and Technology*, volume 37, pages 331–359. Marcel Dekker, Inc., 1997.
- [19] R. Baeza-Yates. Introduction to data structures and algorithms related to information retrieval. In W. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Algorithms and Data Structures*, chapter 2, pages 13–27. Prentice-Hall, 1992. More than **70** citations.
- [20] D. Harman, E. Fox, R. Baeza-Yates, and W. Lee. Inverted files. In W. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Algorithms and Data Structures*, chapter 3, pages 28–43. Prentice-Hall, 1992. More than **420** citations.
- [21] G.H. Gonnet, R. Baeza-Yates, and T. Snider. New indices for text: PAT trees and PAT arrays. In W. Frakes and R. Baeza-Yates, editors, *Information Retrieval: Algorithms and Data Structures*, chapter 5, pages 66–82. Prentice-Hall, 1992. More than **390** citations.

11.3 International Journals

- [1] Tayfun Kucukyilmaz, B. Barla Cambazoglu, Cevdet Aykanat, Ricardo Baeza-Yates. A Machine Learning Approach for Result Caching in Web Search Engines. *Information Processing & Management* 53 (4), July 2017, 834-850.
- [2] Janette Lehmann, Carlos Castillo, Mounia Lalmas, Ricardo Baeza-Yates. Story-focused Reading in Online News and its Potential for User Engagement. *JASIST* 68(4), pp. 869–883, 2017.
- [3] Luz Rello, Ricardo Baeza-Yates. How to Present more Readable Text for People with Dyslexia. *Universal Access in the Information Society* 16(1), pp. 29–49, Springer, 2017.
- [4] Luz Rello, Ricardo Baeza-Yates. The Effect of Font Type on Screen Readability by People with Dyslexia. *ACM Transactions on Accessibility* 8(4), pp. 15–33, 2016.
- [5] David Martin-Borregon, Luca Maria Aiello, Przemyslaw Grabowicz, Alejandro Jaimés, Ricardo Baeza-Yates. Characterization of online groups along space, time, and social dimensions. *EPJ Data Science* 3(1): 1–37, 2014.
- [6] Sandra Alvarez-García, Ricardo Baeza-Yates, Nieves Brisaboa, Josep-Lluis Larriba-Pey, Oscar Pedreira. Automatic Multi-Partite Graph Generation from Arbitrary Data. *Journal of Systems and Software* 94: 72–86, 2014.
- [7] Alexandre Francisco, Ricardo Baeza-Yates, Arlindo Oliveira. Mining query log graphs towards a query folksonomy. *Concurrency and Computation: Practice and Experience* 24(17): 2179–2192, 2012.
- [8] Victor Pascual-Cid, Ricardo Baeza-Yates, Juan Carlos Dürsteler. Visual Web Mining for Website Evaluation. *Journal of Web Engineering* 9(4), 347–368, 2010.
- [9] Barbara Poblete, Myra Spiliopolou, Ricardo Baeza-Yates. Privacy-Preserving Query Log Mining for Business Confidentiality Protection, *ACM Transactions on the Web* 4 (3), 2010.
- [10] Mauricio Marín, Verónica Gil-Costa, Carolina Bonacic, Ricardo Baeza-Yates, Issac D. Scherson. Sync/Async Parallel Search for the Efficient Design and Construction of Web Search Engines, *Parallel Computing* 36(4): 153-168, 2010.
- [11] Diego Puppín, Raffaele Perego, Fabrizio Silvestri, Ricardo Baeza-Yates. Tuning the Capacity of Search Engines: Load-driven Routing and Incremental Caching to Reduce and Balance the Load, *ACM Transactions on Information Systems* 28 (2), April 2010.
- [12] Bernard Jansen, Thomas J. Flaherty, Ricardo Baeza-Yates, Lee Hunter, Brendan Kitts, and Jamie Murphy. The Components and Impact of Sponsored Search, *IEEE Computer* 42 (5), May 2009, 98–101.
- [13] David Nettleton and Ricardo Baeza-Yates. Web Retrieval: Techniques for the Aggregation and Selection of Queries and Answers, *International Journal of Intelligent Systems* 23(12), 1223-1234, 2008.
- [14] Ricardo Baeza-Yates, Aristides Gionis, Flavio Junqueira, Vanessa Murdock, Vassilis Plachouras and Fabrizio Silvestri. Design Trade-Offs for Search Engine Caching, special issue in Query Log Analysis, *ACM Transactions on the Web*, 2(4), Article 4, 2008.

- [15] Luca Becchetti, Carlos Castillo, Debora Donato, Ricardo Baeza-Yates, and Stefano Leonardi. Link Analysis for Web Spam Detection, *ACM Transactions on the Web* 2(1), Article 2, 2008. More than **80** citations.
- [16] Ricardo Baeza-Yates, Peter Mika, and Hugo Zaragoza. Search, Web 2.0, and the Semantic Web. In Trends and Controversies: Near-Term Prospects for Semantic Technologies, R. Benjamins, editor. *IEEE Intelligent Systems* 23 (1), 80–82, Jan-Feb 2008.
- [17] Ricardo Baeza-Yates, Carlos Hurtado and Marcelo Mendoza. Improving Search Engines by Query Clustering, *JASIST* 58(12), 1793–1804, October 2007.
- [18] Alistair Moffat, William Webber, Justin Zobel and Ricardo Baeza-Yates. A Pipelined Architecture for Distributed Text Query Evaluation. *Information Retrieval* 10(3):205-231, June 2007. More than **90** citations.
- [19] Ricardo Baeza-Yates, Carlos Castillo and Efthimis N. Efthimiadis: Characterization of National Web Domains. *ACM Transactions on Internet Technology* 7, Issue 2, 2007. More than **110** citations.
- [20] Ricardo Baeza-Yates and Carlos Castillo. Crawling the Infinite Web: Five Levels are Enough. *Journal of Web Engineering* 6, 49-72, 2007.
- [21] Claudine Badue, Ricardo Baeza-Yates, Berthier Ribeiro-Neto, Artur Ziviani, and Nivio Ziviani. Analyzing Imbalance among Homogeneous Index Servers in a Web Search System. *Information Processing & Management* 43, 592-608, 2007.
- [22] Ricardo Baeza-Yates, Paolo Boldi and Carlos Castillo. Generic Damping Functions for Propagating Importance in Link-Based Ranking, *Journal of Internet Mathematics* 3(4), 445-478, 2006.
- [23] Alexander Jaimes, Javier Ruiz-del-Solar, Rodrigo Verschae, Ricardo Baeza-Yates, Carlos Castillo, Dinko Yaksic, and Emilio Davis. On the Image Content of a Web Segment: Chile as a Case Study, *Journal of Web Engineering* 3 (2), 153–168, 2004.
- [24] Ricardo Baeza-Yates. Information Retrieval in the Web: beyond current search engines, *International Journal on Approximated Reasoning* 34 (2-3), 97–104, 2003. More than **60** citations.
- [25] G. Navarro, R. Baeza-Yates, and J.M.A. Arcoverde. *Matchsimile*: A Flexible Approximate Matching Tool for Searching Proper Names. *JASIST* 54(1): 3-15, 2003.
- [26] R. Baeza-Yates and G. Navarro. XQL and Proximal Nodes (special issue on XML and Information Retrieval). *JASIST* 53(6): 504–514, 2002.
- [27] Gonzalo Navarro, Ricardo Baeza-Yates, Erkki Sutinen and Jorma Tarhio. Indexing Methods for Approximate String Matching (special issue on “Managing Text Natively and in DBMSs”). *IEEE Data Engineering Bulletin* 24(4): 19–27, 2001. More than **200** citations.
- [28] Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates and José Luis Marroquín. Searching in Metric Spaces. *ACM Computing Surveys* 33(3):273–321, 2001. More than **1,250** citations.

- [29] Nivio Ziviani, Edleno de Moura, Gonzalo Navarro and Ricardo Baeza-Yates. Compression: A Key for Next-Generation Text Retrieval Systems. *IEEE Computer* 33(11):37-44 (cover feature), November 2000. More than **140** citations.
- [30] R. Baeza-Yates and G. Navarro. A Hybrid Indexing Method for Approximate String Matching *Journal of Discrete Algorithms (JDA)* 1(1), 205-239, 2000. More than **140** citations.
- [31] Gonzalo Navarro, Edleno Silva de Moura, Marden S. Neubert, Nivio Ziviani, Ricardo Baeza-Yates: Adding Compression to Block Addressing Inverted Indexes. *Information Retrieval* 3(1): 49-77, 2000. More than **110** citations.
- [32] Edleno de Moura, Gonzalo Navarro, Nivio Ziviani and Ricardo Baeza-Yates. Fast and flexible word searching on compressed text. *ACM Transactions on Information Systems (TOIS)* 18(2): 113-139, 2000. More than **320** citations.
- [33] G. Navarro, E. Barbosa, R. Baeza-Yates, W. Cunto, and N. Ziviani, Binary Searching with Non-uniform Costs and Its Application to Text Retrieval. *Algorithmica* 27 (2): 145–169, 2000.
- [34] R. Baeza-Yates and G. Navarro. Block-Addressing Indices for Approximate Text Retrieval. *J. of the American Society for Information Science (JASIS)* 51(1):69–82, Jan 2000. More than **100** citations.
- [35] G. Navarro and R. Baeza-Yates. Very fast and simple approximate string matching. *Information Processing Letters* 72: 65-70. 1999. More than **50** citations.
- [36] R. Baeza-Yates and G. Navarro. Faster approximate string matching. *Algorithmica*, 23 (2):127–158, 1999. More than **180** citations.
- [37] G. Navarro and R. Baeza-Yates. Proximal Nodes: a model to query document databases by content and structure. *ACM TOIS*, 15(4):401–435, Oct 1997. More than **200** citations.
- [38] R. Baeza-Yates and G. Navarro. Integrating content and structure in text retrieval. *ACM SIGMOD Record*, 25(1):67–79, March 1996. More than **150** citations.
- [39] R. Baeza-Yates and G.H. Gonnet. Fast text searching for regular expressions or automaton searching on tries. *Journal of the ACM*, 43(6):915–936, Nov 1996. More than **90** citations.
- [40] R. Baeza-Yates, E. Barbosa, and N. Ziviani. Hierarchies of indices for text searching. *Information Systems*, 21(6):497–514, 1996. **COMPAQ Award to the best Brazilian CS paper for that year.**
- [41] R. Baeza-Yates and C.H. Perleberg. Fast and practical approximate pattern matching. *Information Processing Letters*, 59:21–27, 1996. More than **70** citations.
- [42] R. Baeza-Yates, J. Culberson, and G. Rawlins. Searching in the plane. *Information and Computation*, 106(2):234–252, Oct 1993. More than **370** citations.
- [43] R. Baeza-Yates and G.H. Gonnet. A new approach to text searching. *Communications of the ACM*, 35:74–82, Oct 1992. More than **720** citations.

- [44] U. Manber and R. Baeza-Yates. An algorithm for string matching with a sequence of don't cares. *Information Processing Letters*, 37:133–136, February 1991. More than **100** citations.
- [45] R. Baeza-Yates. Searching Subsequences. *Theoretical Computer Science*, 78:363–376, 1991. More than **90** citations.
- [46] R. Baeza-Yates and P-Å. Larson. Performance of B^+ -trees with partial expansions. *IEEE Trans. on Knowledge and Data Engineering*, 1:248–257, June 1989. More than **50** citations.
- [47] R. Baeza-Yates. Improved string searching. *Software-Practice and Experience*, 19(3):257–271, 1989. More than **90** citations.
- [48] R. Baeza-Yates. Expected behaviour of B^+ -trees under random insertions. *Acta Informatica*, 26(5):439–472, 1989. More than **60** citations.

11.4 International Conferences

- [1] Lorena Recalde, David Nettleton, Ricardo Baeza-Yates, Ludovico Boratto. Detection of Trending Topic Communities: Bridging Content Creators and Distributors. In *Hypertext 2017*, Prague, Check Republic, Jul 2017.
- [2] Liangda Li, Hongbo Deng, Anlei Dong, Yi Chang, Ricardo Baeza-Yates, Hongyuan Zha. Exploring Query Auto-Completion and Click Logs for Contextual-Aware Web Search and Query Suggestion. In *WWW 2017*, pp. 539–548, Perth, Australia, Apr 2017.
- [3] Ioanna Tsalouchidou, Gianmarco De Francisci Morales, Francesco Bonchi, Ricardo Baeza-Yates. Scalable Dynamic Graph Summarization. In *IEEE Big Data 2016*, pp. 1032–1039, Washington D.C., USA, Dec 2016.
- [4] Luca Aiello, Ioannis Arapakis, Ricardo Baeza-Yates, Xiao Bai, Nicola Barbieri, Amin Mantrach, Fabrizio Silvestri. The Role of Relevance in Sponsored Search. In *CIKM 2016*, pp. 185–194, Indianapolis, USA, November 2016.
- [5] Mihajlo Grbovic, Djuric Nemanja, Vladan Radosavljevic, Fabrizio Silvestri, Ricardo Baeza-Yates, Andrew Feng, Erik Ordentlich, Lee Yang, Gawin Owens. Scalable Semantic Matching of Search Queries to Ads in Sponsored Search Advertising. In *SIGIR 2016*, pp. 375–384, Pisa, Italy, July 2016.
- [6] Aston Zhang, Amit Goyal, Ricardo Baeza-Yates, Yi Chang, Jiawei Han, Carl A. Gunter, Hongbo Deng. Towards Mobile Query Auto-Completion: A Mobile Application-Aware Approach. In *WWW 2016*, Montreal, Canada, April 2016.
- [7] Eduardo Graells-Garrido, Mounia Lalmas, Ricardo Baeza-Yates. Encouraging Diversity and Representation Awareness in Geographically Centralized Content. In *ACM Intelligent Users Interfaces (IUI)*, ACM Press, Sonoma, CA, USA, March 2016.
- [8] Eduardo Graells-Garrido, Mounia Lalmas, Ricardo Baeza-Yates. Data Portraits and Intermediary Topics: Encouraging Exploration of Politically Diverse. In *ACM Intelligent Users Interfaces (IUI)*, ACM Press, Sonoma, CA, USA, March 2016.

- [9] Liangda Li, Hongbo Deng, Anlei Dong, Yi Chang, Hongyuan Zha, Ricardo Baeza-Yates. Analyzing User's Sequential Behavior in Query Auto-Completion via Markov Processes. In *ACM SIGIR 2015*, Santiago, Chile, August 2015.
- [10] Ricardo Baeza-Yates, Luz Rello, Julia Dembowski. CASSA: A Context-Aware Synonym Simplification Algorithm. In *NAACL-HLT 2015*, Denver, USA, June 2015.
- [11] Ricardo Baeza-Yates, Paolo Boldi, Flavio Chierichetti. Essential Pages are Easy to Find. In *WWW 2015*, Florence, May 2015. **Nominated to the best paper award.**
- [12] Ricardo Baeza-Yates, Di Jiang, Fabrizio Silvestri, Beverly Harrison. Predicting the Next App that You Are Going to Use. In *WSDM 2015*, Shanghai, Feb 2015.
- [13] Diego Saez-Trumper, Yabing Liu, Ricardo Baeza-Yates, Balachander Krishnamurthy, Alan Milsove. Beyond CPM and CPC: determining the value of users on OSNs. In *COSN'14: Proceedings of the second ACM conference on Online social networks*. ACM, Dublin, Ireland, October 2014.
- [14] Luz Rello, Ricardo Baeza-Yates. Evaluation of Dyswebxia, A Reading App Designed for People with Dyslexia. *W4A 2014*, Seoul, South Korea, Apr 2014. **Best paper runner up.**
- [15] Cigdem Aslay, Nicola Barbieri, Francesco Bonchi, Ricardo Baeza-Yates. Online Topic-Aware Influence Maximization Queries. In *EDBT 2014*, Athens, Greece, 295–306, Mar 2014.
- [16] Guillem Frances, Xiao Bai, B. Barla Cambazoglu, Ricardo Baeza-Yates. Improving the Efficiency of Multi-site Web Search Engines. In *WSDM 2014*, New York, USA, 3–12, Feb 2014.
- [17] Janette Lehmann, Mounia Lalmas, Georges Dupret, Ricardo Baeza-Yates. Online Multitasking - What Web Analytics think we do and what we really do. In *CIKM 2013*, Burlingame, Oct 2013.
- [18] Elad Yom-Tov, Mounia Lalmas, Ricardo Baeza-Yates, Georges Dupret, Janette Lehmann, Pinar Donmez. Measuring Inter-Site Engagement. In *IEEE International Conference on Big Data (BigData 2013)*, Santa Clara, USA, Oct 2013.
- [19] Hossein Vahabi, Margareta Ackerman, David Loker, Ricardo Baeza-Yates, Alejandro López-Ortiz. Orthogonal query recommendation. In *ACM RecSys 2013*, Hong Kong, China, Oct 2013, 33–40.
- [20] Luz Rello, Ricardo Baeza-Yates, Stefan Bott, Horacio Saggion. Simplify or help?: text simplification strategies for people with dyslexia. In *W4A 2013*, Rio de Janeiro, Brazil, May 2013. **Best paper award.**
- [21] Diego Saez-Trumper, Giovanni Comarella, Virgilio Almeida, Ricardo Baeza-Yates, Fabricio Benvenuto. Finding Trendsetters in Information Networks. In *ACM KDD 2012*, Beijing, China, Aug 2012.
- [22] Luz Rello, Ricardo Baeza-Yates, Ruslan Mitkov. Elliphant: Improved Automatic Detection of Zero Subjects and Impersonal Constructions in Spanish. In *EACL 2012*, Avignon, France, Apr 2012, 706–715.
- [23] Ricardo Baeza-Yates, Simon Jonassen. Modeling Static Caching in Web Search Engines. In *ECIR 2012*, Barcelona, Spain, Springer LNCS 7224, Apr 2012, 436–446.

- [24] Cristina González-Caro, Ricardo Baeza-Yates. A Multi-faceted Approach to Query Intent Classification. In *SPIRE 2011*, Pisa, Italy, Oct 2011, Springer LNCS 7024, 368–379.
- [25] Vinay Jethava, Liliana Calderón-Benavides, Chiranjib Bhattacharyya, Devdatt Dubhashi, Ricardo Baeza-Yates. Scalable multi-dimensional user intent identification using tree structured distributions. In *SIGIR 2011*, Beijing, China, Jul 2011.
- [26] Omar Alonso, Ricardo Baeza-Yates. Design and Implementation of Relevance Assessments using Crowdsourcing. In *ECIR 2011*, Dublin, Ireland, Springer LNCS 6611, Apr 2011, 153–164.
- [27] Shuai Ding, Josh Attenberg, Ricardo Baeza-Yates, Torsten Suel. Batch Query Processing for Web Search Engines. In *WSDM 2011*, ACM Press, Hong Kong, Feb 2011, 137–146.
- [28] B. Barla Cambazoglu, Hugo Zaragoza, Ricardo Baeza-Yates. Web Search Solved? All Result Rankings the Same? In *ACM CIKM 2010*, Toronto, Nov 2010, 529–538.
- [29] Alexandre P. Francisco, Ricardo Baeza-Yates, Arlindo L. Oliveira. Mining Large Query Induced Graphs towards a Hierarchical Query Folksonomy. *SPIRE 2010*, Springer LNCS 6393, Los Cabos, Oct 2010, 237–242.
- [30] Omar Alonso, Michael Gertz, Ricardo Baeza-Yates. Temporal Analysis of Document Collections: Framework and Applications. *SPIRE 2010*, Springer LNCS 6393, Los Cabos, Oct 2010, 290–296.
- [31] Ilaria Bornino, Debora Donato, Ricardo Baeza-Yates. Coniunge et Impera: Multiple-graph mining for Query-log analysis. In *Proceedings of PKDD'2010*, Barcelona, Spain, Sept 2010, 168–183.
- [32] B. Barla Cambazoglu, Emre Varol, Enver Kayaaslan, Cevdet Aykanat, Ricardo Baeza-Yates. Query Forwarding in Geographically Distributed Search Engines. In *Proceedings of SIGIR'2010*, Geneva, Switzerland, July 2010, 90–97.
- [33] Ricardo Baeza-Yates, Aristides Gionis, Flavio Junqueira, Vassilis Plachouras, Luca Tello. On the Feasibility of Multi-Site Web Search Engines. In *ACM CIKM 2009*, Hong Kong, China, November 2009, 425-434. **Best paper award.**
- [34] Omar Alonso, Michael Gertz, Ricardo Baeza-Yates. Clustering and Exploring Search Results using Timeline Constructions. In *ACM CIKM 2009*, Hong Kong, China, November 2009.
- [35] Ricardo Baeza-Yates, Christian Middleton, Carlos Castillo. The Geographical Life of Search. In *ACM/IEEE/WIC Web Intelligence*, Milano, Italy, September 2009.
- [36] Ricardo Baeza-Yates, Vanessa Murdock, Claudia Hauff. Efficiency Trade-offs in Two-tier Web Search Systems. In *SIGIR 2009*, Boston, USA, July 2009, 163-170.
- [37] Berkant Barla Cambazoglu, Vassilis Plachouras, Ricardo Baeza-Yates. Quantifying Performance and Quality Gains in Distributed Web Search Engines. In *SIGIR 2009*, Boston, USA, July 2009, 411-418.
- [38] Alvaro Pereira, Ricardo Baeza-Yates, Nivio Ziviani, Jesús Bisbal. A Model for Fast Web Mining Prototyping. In *ACM WSDM 2009*, Barcelona, Feb 2009.

- [39] Claudia Hauff, Vanessa Murdock, Ricardo Baeza-Yates. Improved Query Difficulty Prediction for the Web. In *CIKM 2008*, Napa Valley, USA, Oct 2008.
- [40] Gleb Skobeltsyn, Flavio Junqueira, Vassilis Plachouras, Ricardo Baeza-Yates. ResIn: A Combination of Result Caching and Index Pruning for High-performance Web Search Engines. In *SIGIR 2008*, Singapore, Jul 2008.
- [41] Ricardo Baeza-Yates, Álvaro R. Pereira Jr., Nivio Ziviani. Genealogical trees on the Web: a search engine user perspective. In *WWW 2008*, Beijing, China, Apr 2008, 367-376.
- [42] Barbara Poblete, Ricardo Baeza-Yates. Query-sets: using implicit feedback and query patterns to organize Web documents. In *WWW 2008*, Beijing, China, Apr 2008, 41-50.
- [43] Barbara Poblete, Myra Spiliopoulou, Ricardo Baeza-Yates. Website Privacy Preservation for Query Log Publishing. In *First ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD (PinKDD'07)*, F. Bonchi, E. Ferrari, B. Malin and Y. Saygin (Eds.), Springer LNCS 4890, San Jose, California, USA, 80–96, 2008. **Best paper award.**
- [44] Ricardo Baeza-Yates, Flavio Junqueira, Vassilis Plachouras, Hans F. Witschel. Admission Policies for Caches of Search Engine Results, In *SPIRE 2007*, LNCS 4726, Springer, Santiago, Chile, Oct 2007, 74-85.
- [45] Ricardo Baeza-Yates, Alessandro Tiberi. Extracting Semantic Relations from Query Logs. In *ACM KDD 2007*, San Jose, California, USA, August 2007, 76–85. More than **250** citations.
- [46] Ricardo Baeza-Yates, Aristides Gionis, Flavio Junqueira, Vanessa Murdock, Vassilis Plachouras, Fabrizio Silvestri. The Impact of Caching on Search Engines. In *30th Annual International ACM SIGIR Conference*, Amsterdam, Netherlands, July 2007, 183–190. More than **130** citations.
- [47] Ricardo Baeza-Yates, Georges Dupret, Javier Velasco. A Study of Mobile Search Queries in Japan. Workshop on Query Log Analysis: Social and Technological Challenges, *WWW 2007*, Banff, Canada, May 2007.
- [48] Ricardo Baeza-Yates, Liliana Calderón, Cristina González. The Intention Behind Web Queries. *SPIRE 2006*, LNCS Springer, Glasgow, Scotland, October 2006. More than **140** citations.
- [49] Ricardo Baeza-Yates, Paolo Boldi, Carlos Castillo. Generalizing PageRank: Damping Functions for Link-Based Ranking Algorithms. In *Proceedings of ACM SIGIR 2006*. Seattle, Washington, USA, August 2006. More than **80** citations.
- [50] Luca Becchetti, Carlos Castillo, Debora Donato, Stefano Leonardi, Ricardo Baeza-Yates. Link-Based Characterization and Detection of Web Spam. In *Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*. Seattle, USA, August 2006. More than **130** citations.
- [51] R. Baeza-Yates, C. Castillo, M. Marín, A. Rodríguez. Crawling a Country: Better Strategies than Breadth-First for Page Ordering. In *WWW 2005, Industrial Track*, ACM Press, Chiba, Japan, May 2005. More than **90** citations.

- [52] R. Baeza-Yates, J. Ruiz-del-Solar, R. Verschae, C. Hurtado, C. Castillo. Content-based Image Retrieval and Characterization on Specific Web Collections. In *Conference on Image and Video Retrieval*, LNCS, Springer, Dublin, Ireland, July 2004.
- [53] R. Baeza-Yates. A Fast Set Intersection Algorithm for Sorted Sequences, In *15th Combinatorial Pattern Matching 2004*, LNCS, Springer, Istanbul, Turkey, July 2004. More than **90** citations.
- [54] Ricardo Baeza-Yates, Carlos A. Hurtado, Marcelo Mendoza. Query Recommendation Using Query Logs in Search Engines, Current Trends in Database Technology - EDBT 2004 Workshops, Workshop on Clustering Information over the Web, Heraklion, Crete, Greece, March 14-18, 2004, Revised Selected Papers. LNCS 3268, Springer, p. 588-596. More than **140** citations.
- [55] Fidel Cacheda, Ricardo Baeza-Yates. An Optimistic Model for Searching Web Directories. In *Proceedings of the 26th European Conference on IR Research (ECIR 2004)*. Lecture Notes in Computer Science 2997, Advances in Information Retrieval, Springer, pp 364-377, March 2004.
- [56] R. Baeza-Yates, F. Saint-Jean. A Three Level Search Engine Index based in Query Log Distribution. In *SPIRE 2003*, Springer LNCS, Manaus, Brazil, October 2003.
- [57] R. Baeza-Yates, C. Castillo. Balancing Volume, Quality and Freshness in Web Crawling. In *Hybrid Intelligent Systems 2002*, IOS Press, Santiago, December 2002.
- [58] R. Baeza-Yates, F. Saint-Jean, C. Castillo. Web Dynamics, Age and Page Quality, In *Proceedings of SPIRE 2002*, LNCS, Springer, Lisbon, Portugal, September 2002. More than **90** citations.
- [59] R. Baeza-Yates, C. Castillo, Relating Web Characteristics with Link Based Web Page Ranking, In *Proceedings of SPIRE 2001*, IEEE CS Press, Laguna San Rafael, Chile, pp. 21-32, November 2001. More than **50** citations.
- [60] C. Badue, R. Baeza-Yates, B. Ribeiro-Neto, N. Ziviani. Distributed Query Processing Using Partitioned Inverted Files, In *Proceedings of SPIRE 2001*, IEEE CS Press, Laguna San Rafael, Chile, pp. 10-20, November 2001. More than **120** citations.
- [61] E. Moura, G. Navarro, N. Ziviani, R. Baeza-Yates. Fast searching on compressed text allowing errors. In *21th Annual International Conf. on Research and Development in Information Retrieval (SIGIR'98)*, Melbourne, Australia, 1998, 298-306. More than **60** citations.
- [62] R. Baeza-Yates, G. Navarro. Block-addressing indices for approximate text retrieval. In *Proc. CIKM'97*, pages 1-8, Las Vegas, USA, 1997.
- [63] R. Baeza-Yates, G. Navarro. A faster algorithm for approximate string matching. In Dan Hirschberg and Gene Myers, editors, *Combinatorial Pattern Matching (CPM'96)*, LNCS 1075, pages 1-23, Irvine, CA, Jun 1996. More than **110** citations.
- [64] Eduardo Barbosa, Gonzalo Navarro, Ricardo Baeza-Yates, Chris Perleberg, Nivio Ziviani. Optimized binary search and text retrieval. In *European Symposium on Algorithms*, Corfu, Greece, September 1995.

- [65] Gonzalo Navarro, Ricardo Baeza-Yates. A language for queries on structure and contents of textual databases. In *18th ACM Conference on Research and Development in Information Retrieval (SIGIR'95)*, pages 93–101, Seattle, WA, USA, July 1995. More than **110** citations.
- [66] R. Baeza-Yates, W. Cunto, U. Manber, S. Wu. Proximity matching using fixed queries trees. In M. Crochemore and D. Gusfield, editors, *5th Combinatorial Pattern Matching*, LNCS 807, pages 198–212, Asilomar, CA, June 1994. Springer-Verlag. More than **200** citations.
- [67] R. Baeza-Yates, C.H. Perleberg. Fast and practical approximate pattern matching. In A. Apostolico, M. Crochemore, Z. Galil, and U. Manber, editors, *Combinatorial Pattern Matching*, Springer LNCS 644, pages 185–192, Tucson, AZ, April/May 1992. More than **120** citations.
- [68] R. Baeza-Yates, G.H. Gonnet. A new approach to text searching. In *Proc. of 12th ACM SIGIR*, pages 168–175, Cambridge, Mass., June 1989. (Addendum in ACM SIGIR Forum, V. 23, Numbers 3, 4, 1989, page 7.).
- [69] R. Baeza-Yates, G.H. Gonnet. Efficient text searching of regular expressions. In *ICALP'89*, G. Ausiello, M. Dezani-Ciancaglini, and S. Ronchi Della Rocca, editors, Springer LNCS 372, pages 46–62, Stresa, Italy, July 1989.
- [70] R. Baeza-Yates, J. Culberson, G. Rawlins. Searching with uncertainty. In R. Karlsson and A. Lingas, editors, *Proceedings SWAT 88, First Scandinavian Workshop on Algorithm Theory, Lecture Notes in Computer Science 318*, pages 176–189, Halmstad, Sweden, July 1988. More than **50** citations.

11.5 Selected Invited Papers

- [1] Yannis Kalantidis, Ayman Farahat, Lyndon Kennedy, Ricardo Baeza-Yates, David A. Shamma. Visual Congruent Ads for Image Search. In *ICPR 2016*, pp. 1496-1505, Cancún, Mexico, Dec 2016.
- [2] Ricardo Baeza-Yates. Towards a Distributed Search Engine. In *Algorithms and Complexity (CIAC)*. Tiziana Calamoneri and Josep Diaz, editors. Springer LNCS 6078, Rome, Italy, May 2010, p. 1–5.
- [3] Ricardo Baeza-Yates. The Next Generation of Search. In *Online Information 2009*, London, UK, December 2009, 3-9.
- [4] Ricardo Baeza-Yates, Massi Ciaramita, Peter Mika, and Hugo Zaragoza. Towards Semantic Search. In *NLDB 2008*, E. Kapetanios, V. Sugumaran and M. Spiliopoulou, editors. Springer LNCS 5039, London, UK, June 2008, p. 4-11.
- [5] Ricardo Baeza-Yates, Raghu Ramakrishnan. Data challenges at Yahoo!. In *EDBT 2008*, Nantes, France, Mar 2008, 652-655.
- [6] Ricardo Baeza-Yates, Carlos Castillo, Flavio Junqueira, Vassilis Plachouras and Fabrizio Silvestri. Challenges on Distributed Web Retrieval. In *ICDE 2007*, IEEE Computer Society Press, Estambul, Turkey, 2007, 6–20. More than **80** citations.
- [7] Ricardo Baeza-Yates, Paolo Boldi and José María Gómez. Adversarial Retrieval in the Web. Upgrade, February 2007 (Vol. VIII, issue no. 1), pp. 33-40. In Spanish in *Novática* 185, pp. 29-35. **Best paper of 2007.**

- [8] Ricardo Baeza-Yates. Applications of Web Query Mining. In *European Conference on Information Retrieval (ECIR'05)*, D. Losada, J. Fernández-Luna (editors), Springer LNCS 3408, Santiago de Compostela, Spain, March 2005, 7–22. More than **60** citations.
- [9] Ricardo Baeza-Yates. Challenges in the Interaction of Natural Language Processing and Information Retrieval. *CICLING 2004*, Springer LNAI 2945, Seoul, Korea, 445–456, February 2004.
- [10] Ricardo Baeza-Yates, José Miguel Piquer. Agents, Crawlers and Web Retrieval, Cooperative Information Agents (CIA), LNIA, Springer, Madrid, September 2002.
- [11] R. Baeza-Yates, T. Jones, and G. Rawlins. New Approaches to Information Management: Attribute-Centric Data Systems, 7th International Symposium on String Processing and Information Retrieval (SPIRE), IEEE Computer Science Press, pages 18–27, A Coruña, Spain, 27-29 September, 2000.

11.6 Theses

- [1] R.A. Baeza-Yates. *Efficient Text Searching*. PhD thesis, Dept. of Computer Science, University of Waterloo, May 1989. Also as Research Report CS-89-17. More than **70** citations.
- [2] R.A. Baeza-Yates. Portable graphical editor for three dimensional objects. Master's thesis, Dept. of Electrical Engineering, University of Chile, Santiago, Chile, April 1986. (in Spanish).
- [3] R.A. Baeza-Yates. Analysis of algorithms in search trees. Master's thesis, Dept. of Computer Science, University of Chile, Santiago, Chile, January 1985. (in Spanish).

12 Selected PhD Theses Supervised

Notice the positive gender discrimination with 48% female Ph.D. graduates and 27% female in the rest.

1. Maria Rauschenberger. Early Screening of Dyslexia using a Language-Independent Content Game and Machine Learning. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, October 2019. Co-advisor: Luz Rello.
2. Necati Bora Edizel. Word Embeddings with Applications to Web Search and Advertising. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, July 2019. Co-advisor: Amin Mantrach.
3. Ioanna Tsalouchidou. Temporal Analysis of Dynamic Large Graphs. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, October 2018. Co-advisor: Francesco Bonchi.
4. Lorena Recalde. Modeling User Preferences in Online Social Networks. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, October 2018. Co-advisor: David Nettleton.
5. Çigdem Aslay. From Viral Marketing to Social Advertising: Ad Allocation under Social Influence. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, December 2016. Co-advisor: Francesco Bonchi.

6. Yasir Mehmood. Modeling the Roles of Users and Groups in the Dynamics of Information Propagation. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, April 2016. Co-advisor: Francesco Bonchi.
7. Eduardo Graells-Garrido. Biased Behavior in Web Activities: From Understanding to Unbiased Visual Exploration. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, May 2015. Co-advisor: Mounia Lalmas.
8. Janette Lehmann. From Site to Inter-site User Engagement: Fundamentals and Applications. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, February 2015. Co-advisor: Mounia Lalmas.
9. Ruth García-Gavilanes. User Behavior in Microblogging with a Cultural Emphasis. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, February 2015.
10. Michele Trevisiol. Exploiting Implicit User Activity for Media Recommendation. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, October 2014. Co-advisor: Alejandro Jaimes.
11. Luca Chiarandini. Characterizing and Modelling Web Sessions with Applications. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, October 2014. Co-advisor: Alejandro Jaimes.
12. Diego Saéz-Trumper. Finding Relevant People in Online Social Networks. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, December 2013.
13. Hossein Vahabi. Recommendation Techniques for Web Search and Social Media. Ph.D. in Computer Science, Institute for Advanced Studies, Lucca, Italy. July 2012. Co-advisor: Fabrizio Silvestri.
14. Liliana Calderón-Benavides. Unsupervised Identification of the User's Query Intent in Web Search. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, September 2011.
15. Cristina González-Caro. Supervised Identification of the User Intent of Web Search Queries. Ph.D. in Computer Science, Univ. Pompeu Fabra, Barcelona, Spain, September 2011.
16. Ximena Olivares. Large Scale Image Retrieval based on User Generated Content. Ph.D. in Computer Science, Univ. Pompeu Fabra, Spain, March 2011. Co-advisor: Roelof Van Zwol.
17. Víctor Pascual. Visual Exploration of Web Spaces. Ph.D. in Computer Science, Univ. Pompeu Fabra, Spain, December 2010. Co-advisor: Juan Carlos Dürstelner.
18. Ilaria Bordino. Graph Mining and Its Application to the Web. Ph.D. in Computer Science, Università di Roma "La Sapienza", Italy, and Univ. Pompeu Fabra, Spain, June 2010. Co-advisor: Stefano Leonardi.
19. Barbara Poblete. Query Based Data Mining for the Web. Ph.D. in Computer Science, Univ. Pompeu Fabra, Spain, October 2009. Co-advisor: Myra Spiliopoulou.
20. Omar Alonso. Temporal Information Retrieval. Ph.D. in Computer Science, Univ. of California at Davis, USA, September 2008. Co-advisor: Michael Gertz.

21. Alvaro Pereira Jr. A Model for Fast Web Mining Prototyping: Design, Algebra, Implementation and Use Cases. Ph.D. in Computer Science, Univ. Federal de Minas Gerais, Brazil, September 2008. Co-advisor: Nivio Ziviani.
22. Mendoza, Marcelo. Query Log Mining in Search Engines. Ph.D. in Computer Science, Univ. de Chile, June 2007. Co-advisor: Carlos Hurtado.
23. Badue, Claudine. Design and Analysis of Web Search Systems. Ph.D. in Computer Science, Univ. Fed. de Minas Gerais, Brazil, February 2007. Main advisor: Nivio Ziviani.
24. Castillo, Carlos. Effective Web Crawling, Ph.D. in Computer Science, Univ. de Chile, November 2004.
25. Chávez, Edgar. Similarity Searching on Metric Spaces, Ph.D. in Computer Science, CIMAT, México, September 1999. Co-advisor: José Luis Marroquín.
26. Navarro, Gonzalo. Approximate Text Searching, Ph.D. in Computer Science, Univ. de Chile, December 1998.
27. Barbosa, Eduardo. Efficient Text Retrieval Methods for Secondary Memory, Ph.D. in Computer Science, Univ. Fed. de Minas Gerais, Brazil, July 1995. Main advisor: Nivio Ziviani.

13 Editorial Boards

I am editor of the following journals: Information Systems (Elsevier, Netherlands) since 2002, Computing (Springer, Germany) since 2013, Journal of Web Engineering (USA) since 2002, Theoretical Informatics and Applications - RAIRO (France) since 1998. Grammars (Netherlands) since 2004, Journal of the SBC (Brazil) since 2002, First Monday (USA) since 2001, and El Profesional de la Información (Spain) since 2003.

I was the first Editor-in-Chief of the CLEI Electronic Journal at www.clei.cl from 1998 to 2005. I also was member of the editorial boards of ACM Transactions on Information Systems (ACM, USA) from 2002 to 2014, Information Processing & Management (Elsevier, Netherlands) from 2005 to 2011 (since 2007 as Associated Editor), VLDB Journal (Springer, Germany) from 2001 to 2007, Computer Science Review (Elsevier, Netherlands) from 2013 to 2014, Information Technology for Development (Netherlands), International Journal of Web Information Systems (UK), Journal of Digital Management (India-South Korea) from 2003 to 2005, Computación y Sistemas (México), Novática (Spain), Revista de la SCCC (Chile) and Revista Colombiana de Informática (Colombia).

14 Selected Keynotes

- Bias in the Web, Gödel Annual Seminar, Wien, Austria, Jun 2017. YouTube: <https://www.youtube.com/watch?v=CguWbbQerUM>.
- “Data Monopoly: A New Power?” (in Spanish), The Congress of the Future, Santiago, Chile, Jan 2016. YouTube: <https://www.youtube.com/watch?v=fFlkOKUa2M0>.

- “Does Big Data Empowers the Web?” (in Spanish), TEDx Barcelona, July 2015. YouTube: <https://www.youtube.com/watch?v=JvO1P8yGVqA> (more than **4,000** views).
- Invited speaker or tutorial teacher at many conferences including ACM SIGIR, WWW, ACM WSDM, VLDB, ACM CIKM, ACM Recsys, SDM, ACM Hypertext, ACM Web Science, ECDL, ICDE, ECIR, ECML/PKDD, ESWC, UMAP, ICWE, EDBT, ISSI, SPIRE, CIAC, CLEF, NLDB, SAMT, SSDBM, SISAP, ICTIR, DSAI, ICWI, WISE, CRWIG, Simultools, SRDS, IDA, LA-WEB, etc.; as well as in many workshops.
- Invited speaker in many professional events related to the Web, search technology, and innovation in several countries: Brazil, Chile, Finland, Italy, Mexico, Netherlands, Spain, Switzerland, United Kingdom, and USA.

15 Service at Conferences

- Member of the steering committees of ACM CIKM (since 2014), ICWSM (since 2014), SPIRE since 1998 (co-founder), AMW since 2006 (co-founder), OAIR since 2011, ESSIR since 2009, and LA-WEB since 2003 (founder). I was part of the steering committee of ACM WSDM from 2008 to 2017 (co-founder) and LATIN from April 1995 to March 2006.
- Conference co-chair of The Web Conference 2019 (formerly WWW), San Francisco, USA, May 2019; Conference co-chair of IEEE Big Data 2017, Boston, USA, Dec 2017; Co-organizing chair of ESSIR 2017, Barcelona, Spain, Sep 2017; Conference co-chair of IEEE ICDM 2016, Barcelona, Spain, Dec 2016; Conference chair of ACM SIGIR 2015, Santiago, Chile, Aug 2015; Conference chair of LA-WEB 2014, Ouro Preto, Brazil, Oct 2014; Conference chair of AMW 2014, Cartagena de Indias, Colombia, Jun 2014; Conference chair of LA-WEB 2012, Cartagena de Indias, Colombia, Oct 2012; Conference chair of ECIR 2012, Barcelona, Spain, Apr 2012; Co-organizing chair of ESSIR 2009, Padova, Italy, Sep 2009; Conference chair of ACM WSDM 2009, Barcelona, Spain, Feb 2009; Conference chair of LA-WEB 2008, Vila Velha, Brazil, Oct 2008; Conference chair of SPIRE 2007 and LA-WEB 2007 (collocated) Santiago, Chile, Oct 2007; Conference co-chair of ACM SIGIR 2015 in Bahia, Brazil, Aug 2005; Conference chair of the first Latin American Web conference (LA-WEB 2003) in Santiago, Chile, Nov 2003; Conference co-chair of IEEE PROMSMmNet'97, Santiago, Chile, Nov 1997; Conference co-chair of LATIN'95, Viña del Mar, Chile, Apr 1995; Conference chair of the 2nd South American Workshop on String Processing, Viña del Mar, Chile, Apr 1995; and Conference chair of the XI International Conference of the Chilean Computer Science Society, Santiago, Chile, Oct 1991.
- Panels Co-chair for The Web Conference 2021 (Ljubjana, Slovenia, May 2021); Applied Research co-chair for CIKM 2020 (Galway, Ireland, Oct 2020); Tutorials co-chair for ASONAM 2018 (Barcelona, Spain, Sep 2018); Tutorials co-chair for WWW 2015 (Florence, Italy, May 2015); Workshops co-chair for WWW 2014 (Seoul, South Korea, Apr 2014); Program committee co-chair for WWW 2013 (Rio de Janeiro, Brazil, May 2013); Industry and EU projects tracks co-chair for WWW 2012 (Lyon, France, April 2012); Program committee co-chair for SIGIR 2011 (Beijing, China, Jul 2011); Program committee co-chair for ICWSM 2011 (Barcelona, Spain, Jul 2011); Tutorial chair for WWW 2010 (Raleigh, USA, Apr 2010); Program committee chair for ACM Web Intelligence & Internet

Agent Technology (Milano, Italy, September 2009); Industrial track PC co-chair for KDD 2009 (Paris, France, June 2009); Ibero-American track co-chair for WWW 2009 (Madrid, April 2009); Workshops co-chair for WWW 2008 (Beijing, China, Apr 2008); Program committee vice-chair for ICDM 2008 (Pisa, Italy, December 2008); Program committee co-chair for CIKM 2007 (Lisbon, Portugal, November 2007); Program committee co-chair for the Mexican Computer Science Conference (ENC 2004, Colima, Mexico, September 2004); Program committee co-chair for the First Latin-American Web Conference (LA-WEB, Santiago, Chile, November 2003); Program committee co-chair for the Combinatorial Pattern Matching (CPM, Morelia, Mexico, June 2003); Program committee chair for the Americas of the ACM Conference in Information Retrieval (SIGIR, Tampere, Finland, Aug 2002); Program committee co-chair of IFIP TCS'2002 (Montreal, Canada, Aug 2002); Program committee chair of the XXI Int. Conf. of the Chilean Computer Science Society (Punta Arenas, Chile, Nov 2001); Program committee chair of the XVII Int. Conf. of the Chilean Computer Science Society and the Fourth South American Workshop on String Processing (Valparaíso, Chile, Nov 1997); Program co-chair of LATIN'95, Viña del Mar, April, 1995. Program committee chair of the XIII Int. Conf. of the Chilean Computer Science Society (La Serena, October 1993), and the 1st South American Workshop on String Processing (Belo Horizonte, Brazil, September 1993).

- Founder and organizer of the Future of Web Search workshop: Barcelona (2006), Bertinoro (2007), Andorra (2008) and Ibiza (2009), which continued as the Future of the Social Web workshop (Barcelona, 2011).
- Co-organizer of the first two XML Retrieval workshops at SIGIR in 2000 and 2002. Co-organizer of the first workshop of Content Analysis in the Web 2.0 at WWW 2009. Co-organizer of the Temporal Web workshops at WWW since 2011. Co-organizer of the first CrowdSearch workshop at WWW 2012.

16 Other Relevant Information

- Native language: Spanish. I am fluent in English and I understand Portuguese and Catalan. I can read Italian and French.
- I love geography, traveling, good food and wine, science fiction, and chess. I like to practice tennis, squash and football (soccer).