OMB No. 0925-0001 and 0925-0002 (Rev. 09/17 Approved Through 03/31/2020)

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: R. Baeza-Yates

eRA COMMONS USER NAME (credential, e.g., agency login): N/A

POSITION TITLE: Professor at Northeastern University, Silicon Valley Campus

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE	Completion Date	FIELD OF STUDY
Universidad de Chile, Santiago, Chile	Electronical Engineer	1984	Engineering
Universidad de Chile, Santiago, Chile	M.Sc.	1985	Computer Science
Universidad de Chile, Santiago, Chile	M.Eng. E.E.	1986	Engineering
University of Waterloo, Ontario, Canada	Ph.D.	1989	Computer Science

A. Personal Statement

My main interests are algorithms, data mining and machine learning, which today are part of the core of computer science and data science. Regarding algorithms I have done research in generic search, particularly in text and DNA sequences, as well as web search, which belongs to the field of information retrieval. I also have designed algorithms related to computational biology, object similarity, and computational geometry, among others. Regarding data mining we have studied the evolution of the Web and particularly social media as well as pioneered the subfield of query mining. Regarding machine learning I am interested in efficiency-quality trade-offs as well as the impact of bias on applications that affect people. I have also worked on how to use computer science to help people with dyslexia, including reading comprehension. My research in these areas have crystalized in the following selected contributions: (1) a class of text searching algorithms, bit parallelism, that use bitwise operations to be flexible and run faster; (2) the first optimal algorithm to discover where an unexplored wall ends, which now is used in robotics; (3) novel applications of mining queries to improve search engines; (4) a new distributed architecture for web search; (5) text simplification methods to improve the reading comprehension of people with dyslexia; and (6) characterizing the vicious cycle of bias on the Web.

B. Positions and Honors

1989-1995 Associate Professor, Dept. of Computer Science, Universidad de Chile, Santiago, Chile
1995-2004 Professor, Dept. of Computer Science, Universidad de Chile, Santiago, Chile
2004-2005 ICREA Research Professor, DTIC, Universitat Pompeu Fabra, Barcelona, Spain
2006-2014 VP of Research, Yahoo Labs, Barcelona, Spain
2014-2016 VP of Research & Chief Research Scientist, Yahoo Labs, California, USA
2016-2020 Chief Technology Officer, NTENT, California, USA

Selected Awards and honors

1985 Award of Institute of Engineers of Chile, to the best engineering student graduated in Chile
1993 Award of the Organization of American States for young researchers in Exact Sciences
1994 Recipient of a United States Information Service Important Visitor program, USA
1994 Award of Institute of Engineers of Chile to the best research on engineering on the last 4 years
2003 Named Corresponding Member of the Chilean Academy of Sciences, Chile
2003 Chilean Fulbright Commission Fellowship in Information Technology, USA

2007	J.W. Graham Medal in Computing & Innovation given to distinguished alumni of the University of Waterloo, Canada
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	Named ACM Fellow – 1% of members (ACM is the main USA Computer Science Association)
2010	Founding Member of the Chilean Academy of Engineering, Chile
2010	National Award of the Chilean Association of Engineers, Chile
2011	Named IEEE Fellow for contributions to the development of computer science, USA
2012	ASIS&T Best Information Science Book Award for the 2nd edition of Modern Information Retrieval
2018	National Award for applied research and technology transfer, Scientific Computing Societies of Spain & BBVA Foundation, Spain
2019	Person of the Year, Catalan Association of Telecommunications Engineers with the cooperation of the Association of Computing Engineers of Catalonia, Spain

Selected scientific contributions and publications

1. **Searching Algorithms:** I have designed several well-known algorithms for matching exact and approximated strings in text, as well as new architectures for distributed web search, indexes for semi-structured retrieval, optimization of two-level inverted indexes, etc. These algorithms are being used in the main search engines, main file search tools of operating systems as well as in genomic alignment software such as BLAST.

1. R. Baeza-Yates and G.H. Gonnet. A new approach to text searching. *Communications of the ACM*, 35:74–82, Oct 1992.
2. 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.
3. R. Baeza-Yates and G. Navarro. Faster approximate string matching. *Algorithmica*, 23 (2):127–158, 1999.
4. R. Baeza-Yates, A. Gionis, F. Junqueira, V. Plachouras, L. Telloli. On the Feasibility of Multi-Site Web Search Engines. In ACM CIKM 2009, Hong Kong, China, 425-434, Nov 2009. **Best paper award**.

2. Data Mining: We have pioneered the use of web content to understand the impact of search engines as well as the evolution of social media.

1. R. Baeza-Yates, A. Pereira Jr., N. Ziviani. Genealogical trees on the Web: a search engine user perspective. In WWW 2008, Beijing, China, Apr 2008, 367-376.
2. D. Saez-Trumper, G. Comarela, V. Almeida, R. Baeza-Yates, F. Benevenuto. Finding Trendsetters in Information Networks. In ACM KDD 2012, Beijing, China, Aug 2012.
3. R. Baeza-Yates, P. Boldi, F. Chierichetti. Essential Pages are Easy to Find. In WWW 2015, Florence, May 2015.
4. R. Baeza-Yates, D. Saéz-Trumper. Wisdom of the Crowd or Wisdom of a Few?: An Analysis of Users' Content Generation. In ACM Hypertext & Social Media, pp. 69-74, Cyprus, Sep 2015.

3. Machine Learning and Applications: I have designed innovative applications of machine learning to web advertising, web search caching, query difficulty, query intention, app prediction, etc. as well as the impact of bias on web search, recommender systems and the Web om general.

1. L. Rello, R. Baeza-Yates, Stefan Bott, H. Saggion. Simplify or help? Text simplification strategies for people with dyslexia. In *W4A 2013*, Rio de Janeiro, Brazil, May 2013. **Best paper award.**
2. R. Baeza-Yates, D. Jiang, F. Silvestri, B. Harrison. Predicting the Next App that You Are Going to Use. In WSDM 2015, Shanghai, Feb 2015.
3. R. Baeza-Yates, Z. Liaghat. Quality-efficiency trade-offs in machine learning for text processing. In IEEE Big Data 2017, Boston, USA, Dec 2017.
4. Ricardo Baeza-Yates. Bias on the Web. Communications of ACM 61(6), pp. 54–61, 2018.

4. Computational Biology: We have designed advanced algorithms for DNA matching and we have participated in the sequencing of the DNA of a specific type of peach.

1. R. Baeza-Yates, G. Gonnet. A Fast Algorithm on Average for All-Against-All Sequence Matching. In 6th Symposium on String Processing and Information Retrieval (SPIRE'99), Cancun, Mexico, Sep 1999.
2. H. Silva, R. Baeza-Yates, V. Cambiazo, R. Campos, M. González, L. Meisel, A. Orellana, J. Retamales. Chilean functional genomics in Prunus Persica: An approach towards understanding post- harvest problems in peaches and nectarines. Plant & Animal Genomes XII Conference, San Diego, CA, USA, Jan 2004.
3. R. Campos-Vargas, O. Becerra, R. Baeza-Yates, V. Cambiazo, M. González, L. Meisel, A. Orellana, J. Retamales, H. Silva, and B. G. Defilippi: Seasonal Variation in the Development of Chilling Injury in 'O'Henry' Peaches, Scientia Horticulturae 110: 79-83, 2006.
4. R. Nilo, C. Saffie, K. Lilley, R. Baeza-Yates, V. Cambiazo, R. Campos-Vargas, M. Gonzalez, L.A. Meisel, J. Retamales, H. Silva and A. Orellana. Proteomic analysis of peach fruit mesocarp softening and chilling injury using difference gel electrophoresis (DIGE). *BMC Genomics*, 11–43, 2010

5. Selected Books:

1. 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).
2. R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval: the concepts and technology behind search, second edition. Addison-Wesley, UK, 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). The first edition was published by ACM Press/Addison-Wesley, UK, 513 pages, in 1999, and was translated to Chinese and Korean and printed in special editions for China and India. **This is the most cited book in information retrieval**.
3. B.B. Cambazoglu and R. Baeza-Yates. Scalability Challenges in Web Search Engines, in Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, USA, 122 pages, 2015.

Other publications (more than 500) in
https://scholar.google.com/citations?user=v9xULZwAAAAJ&hl=en&oi=ao

D. Research Support:

Completed (last 5 years):

Ministry of Science and Innovation, Spain (Baeza-Yates co-PI) 01/01/2015-12/31/2015
Privacy Alerts: Would be possible to empower online social network users with tools such that they control the privacy of their data? Budget: 30,000 €

Ministry of Science and Innovation, Spain (Baeza-Yates PI) 01/01/2013-12/31/2015
Understanding Social Media: An Integrated Data Mining Approach. Budget: 200,000 €

Patents

1. R. 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. R. 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. R. Baeza-Yates, A. Gionis, F. Junqueira and V. Plachouras. System and Method for Caching Posting Lists, Yahoo! Inc., October 2007. USA Patent #7,890,488, granted Feb 5, 2011.
4. R. Baeza-Yates, A. Tiberi. Extracting Semantic Relations from Query Logs, Yahoo! Inc., December 2007. USA Patent #7,895,235, granted Feb 22, 2011.
5. L. Barbosa, F. Junqueira, V. Plachouras, R. 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. B. Poblete, R. Baeza-Yates. Classifying Documents using Implicit Feedback and Query Patterns, Yahoo! Inc., July 2008. USA patent #8,645,369, granted February 4, 2014.
7. R. Baeza-Yates, A. Gionis, F. Junqueira, V. Plachouras and L. Telloli. System and Methodology for a Multi-Site Search Engine, Yahoo! Inc., October 2008. USA patent #8,095,545, granted January 10, 2012.
8. B. Poblete, R. Baeza-Yates. Methods for Web Site Analysis, Yahoo! Inc., April 2010. USA patent #8,751,632, granted June 10, 2014.
9. F. Silvestri, R. Baeza-Yates, B. Harrison, D. 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. M. Chandrasekharan, K. Horiguchi, A. Stent, R. Baeza-Yates, Jeffrey Kuwano, Achint Thomas, Yi Chang. Audio Verification, Yahoo! Inc., September 2015. USA Patent\#10277581B2, granted April 30, 2019.
11. L. Zhang, H. Deng, A. Goyal, Y. Chang, R. Baeza-Yates. Computerized System and Method for Search Query Auto-completion, Yahoo! Inc, September 2015. USA Patent #10049149B2, granted August 14, 2018.
12. S. Osindero, I. Kalantidis, D. Shamma, L. Kennedy, A. Farahat, G. Pesavento, R. Baeza-Yates. Method and System for Selecting Supplemental Content using Visual Appearance, Yahoo! Inc, April 2016.