

Antonio Gulli

CONTACT INFORMATION	London, UK Pisa, Italy	<i>E-mail:</i> gulli@di.unipi.it
CITIZENSHIP	Italian, married, three kids	
CURRENT POSITION	Research Scientist in Web Information Retrieval, working at Microsoft as Principal, Development Manager for Bing, and coordinating all the Development activities in London, U.K, Paris, France and Wroclaw, Poland.	
RESEARCH INTERESTS	Web Search especially focused on Freshness and Universal Ranking, Image Search, News Search, Clustering and Classification of dynamic collections, Query processing (auto-suggest, related search, deep links), Query Rewriting (spelling and alteration) Machine Learning on Massive dataset, Mobile and Social Web. My carrier path is about mixing industry and academic work. Nowadays, you must have a great combination of research skills and a just-get-it-done attitude.	
MAIN SEARCH ACHIEVEMENTS	2011- Microsoft.com , Coordinating all the Dev efforts in London, UK, Paris, France and Wroclaw, Poland. Contributing technologies for Bing U.S. and other global markets. 2010 Microsoft.com Principal Development Manager, contributing technologies for 15 International Markets and some features of U.S. Market 2009 Ask.com Global Web Search Ranking (Web Freshness, News, Blog, Video) 2008 DailyBeast News Search Engine & Ask.com BigNews Search engine CTO, Europe and UK for Ask.com 2007 Ask.com News Image Search Engine 2005 GM in the European R&D office, Director Advanced Search Projects for Ask.com Ask.com Blog Search Engine & Ask.com Image Search Engine 2005 Estimated the size of the Web [more than 11.5 billion of Web pages]. Presented 5 search papers at 14th International World Wide Web Conference [2 full, 3 short] 1999-2002 Co-Created one of the first search company in Europe, sold to the second largest European ISP for 7 million Euros	

AWARDS

Microsoft, 2011 Golden start for machine learning pipelines contribution

Time, 2010 50 Best Websites News & Info (assigned to DailyBeast)

IAC, 2007 Emerging Leader

IAC, 2006 Horizon Award for being a top performer above and beyond the expectation

SEARCH

EXPERIENCE

I have been working on Search Engines since 1997. These are some of the projects I have run or contributed:

Current

Microsoft.com, Bing

- "Interestingly Bing's news aggregation service has zoomed past Google's, at least in the UK, according to the survey of 50,000. It's had a growth spurt of 141 percent since May last year, while Google News' readership has fallen by 56 percent" [techeye](#)
- "Bing was launched under Beta tag in the UK. Microsoft at that time promised to remove the tag only under one condition i.e if its experience would be different from the competition and if the results would be outperforming in terms of UK relevancy. The Bing team reached its objective on November 12, 2009 and the credit goes to London-based Search Technology Center. Microsoft says that 60 engineers behind the project in Soho have done extensive job at localizing the Bing global experience for the UK users in just 5 months." [searchnewz](#).
- "No reason to use Google anymore": Bing moves out of beta" [PC&Tech](#)

Past

Ask.com Global Web Search Ranking (Freshness, News/Blog, Video)

- "I looked at search results for Bruce Springsteen in Google, Yahoo, and Ask, to see how current the information was from each. All three had fairly up-to-date news information. Ask actually had a richer results than the other two in terms of the images that they displayed, and their listing of future events or concerts for Springsteen. ... is presently showing a rich mix of images, scheduled events, web pages, and news results for the performer, and if you hover your mouse over the image next to the news results for Bruce Springsteen at Ask.com, the news video starts playing." [seobythesea](#).

Tina Brown's DailyBeast News Engine

- "Five who are changing the face of the Internet" [newsweek](#)
- "Time 2010, 50 Best Websites 2010" [time](#)

Ask.com BigNews (video) news, blog search engine, patent pending

- "Ask.com has a pretty original approach to the old-time, old-school, traditional maybe, view of news." [techcrunch](#)

- "And with the surprising usefulness of the site compared to some of the other social news attempts, I think it's going to succeed in gaining a good core audience.", [rev2](#)
- "My vote - thumbs up. Ask.com has done well integrating social media with news in a clean, easy-to-use interface. Features like story history could save me time", [marketingpilgrim](#)
- "Sounds a bit like PageRank in that it weighs documents by their mentions, but focuses on the leading edge of the Web, on recent documents and new links between documents", [Greg Linden](#)
- "it's interesting that they've managed to integrate the video playing right into the main page since I doubt all the source videos are the same format (Flash, aspx, etc.)", [user comment](#)
- "Any innovation or changes in the news space is refreshing and welcome, and I think Ask BigNews does really well in some area, especially search (no coincidence!)", [user comment](#)

Ask.com in-house Blog and Feed search engine.

- * "Overall, I'm quite impressed with Ask's new blog search engine. I LOVE how it's all mixed up with the best of Bloglines; I think that this will be absolutely one of the most useful blog search engines out there because of this fact alone.", [about.com](#)

Ask.com in-house Fresh Image search engine, patent pending

- * Returned very accurate image results, and it has a well-designed search results page. This is a worthy alternative to Google., [PC-World test 2007](#)
- * "Ask new image search is a step ahead in a notoriously tricky area. With the quality of its image search results, combined with the new Zoom query refinement feature, I'll be using it as my default image search service going forward", [SearchEngineWatch](#)

Ph.D. Thesis: SnakeT Web Clustering Engine, ComeToMyHead News Search Engine, presented at [www05 conference](#) (demo)

Video Search Engine, Image Search Engine, used by Tiscali, Jumpy, CiaoWeb

MP3 Audio Search Engine, used by Tiscali, Jumpy, CiaoWeb, Arianna

Chat & Search Engine on IRC, used by Tiscali, EdisonTel

Shopping Comparison Engine, used by Tiscali, SuperEva

Usenet Search Engine, used by Arianna

Pseudo Real time Usenet news engine, used by Arianna, Corriere della Sera, La Repubblica

Web Search Engine, used by Tiscali, SuperEva, Interfree, Infocamere

Automatic Directory, used by Arianna, Fireball

Spidering system, used by Arianna

Many of these services are no longer online, but you can see the [press releases](#) and some snapshots from the [Waybackmachine](#) in my [LinkedIn profile](#)

SEARCH PATENTS

Similarity detection and clustering of images, assigned
Method and system to present video content, assigned
Sampling Internet user traffic to improve search results, assigned
System and method for categorizing answers such as URLs, pending
Systems and methods for visually selecting information, pending
Systems and methods for clustering information, pending
Systems and methods for personalizing a newspaper, pending
Systems and methods for selecting and organizing information using temporal clustering, pending
System and method for monitoring evolution over time of temporal content, pending
Method and system to present a preview of video content, pending
Method and system to provide targeted advertising with search results, pending

SEOBYTHESEA talking about them:

Ask Since Jeeves: Adding Video News within Search Results to Image and News Clustering
Calculating Search Rankings with User Web Traffic Data
Ask.com on Trends, Freshness, Personalization, and Better Search Results

MANAGEMENT EXPERIENCE

2009-

Principal Development Leader, Microsoft.com for Bing, Leading all the London, Paris and Wroclaw Developers.

2008-2009

CTO, Ask.com for Europe and UK. GM in Italy for Ask.com

2005-2007

Director, Ask.com Advanced Search Products

1999-2002

CEO and Founder, Ideare S.p.A., sold to Tiscali Group

Proven ability to attract great engineering talents and to the group and built strong teams of developers. Proven ability to coordinate and to interact with development centers in different geographical places including U.S. (S.Francisco, NYC, Seattle), Europe (London, Hamburg, Munich, Paris, Pisa, Wroclaw), China (Beijing, Hangzhou), India (Hyderabad). Proven ability to drive fast development and to ship multiple products per year in many different markets and to support large scale rapid growth. Proven ability to work in different contexts, ranging from small and dynamic startups to large corporations. More information available on my LinkedIn Profile

Book

A. Gulli. *Clustering and Ranking for Web Information Retrieval*, VDM Verlag Dr. Mueller, ISBN: 978-3836456579, 2008

Journals

P. Ferragina and A. Gulli. A personalized search engine based on web-snippet hierarchical clustering. *Journal of Software: Practice and Experience*, 2008, volume 2

G. M. Del Corso, A. Gulli, and F. Romani. Efficient Sparse Linear System Solution of the PageRank. *The journal of ATI (Asociacin de Tcnicos de Informtica)*. Invited paper, 2007

Ferragina and A. Gulli. SnakeT: A personalized search-result clustering engine, *The journal of ATI (Asociacin de Tcnicos de Informtica)*. Invited paper, 2007

G. M. Del Corso, A. Gulli, and F. Romani. Comparison of Krylov Subspace Methods on the PageRank Problem. *Journal of Computational and Applied Mathematics*, 2006

G. M. Del Corso, A. Gulli, and F. Romani. Fast pagerank computation via a sparse linear system. *Journal of Internet Mathematics*, 2005, volume 2, number 3

Proceedings

A. Gulli, S. Cataudella, L. Foschini TC-SocialRank: Ranking the Social Web, *Workshop on Algorithms and Models for the Web-Graph - WAW*, 2009

A. Signorini, A. Gulli, A. Segre - Distributed Marketplaces using P2P Networks and Public-Key Cryptography. In *Proceedings of 3rd International ICST Conference on Scalable Information Systems*, 2008

G. M. Del Corso, A. Gulli, and F. Romani. Ranking a stream of news. In *Proceedings of 14th International World Wide Web Conference*, pages 97-106, Chiba, Japan, 2005.

A. Gulli. The anatomy of a news search engine. In *Proceedings of 14th International World Wide Web Conference*, pages 880-881, Chiba, Japan, 2005.

A. Gulli and A. Signorini. Building an open source meta search engine. In *Proceedings of 14th International World Wide Web Conference*, pages 1004-1005, Chiba, Japan, 2005.

A. Gulli and A. Signorini. The indexable web is more than 11.5 billion pages. In *Proceedings of 14th International World Wide Web Conference*, pages 902-903, Chiba, Japan, 2005.

P. Ferragina and A. Gulli. A personalized search engine based on web-snippet hierarchical clustering. In *Proceedings of 14th International World Wide Web Conference*, pages 801-810, Chiba, Japan, 2005.

P. Ferragina and A. Gulli. The anatomy of a hierarchical clustering engine for web-page, news and book snippets. In *Proceedings of the 4th IEEE International Conference on Data Mining*, pages 395-398, Brighton, UK, 2004.

P. Ferragina and A. Gulli. The anatomy of snaket: A hierarchical clustering engine for web-page snippets. In Proceedings of 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, pages 506-508, Pisa, Italy, 2004.

G. M. Del Corso, A. Gulli, and F. Romani. Fast pagerank computation via a sparse linear system (extended abstract). In Proceedings of Algorithms and Models for the Web-Graph: Third International Workshop, WAW, pages 118-130, Rome, Italy, 2004

P. Ferragina and A. Gulli. Experimenting snaket: A hierarchical clustering engine for webpage snippets. In Proceedings of 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, pages 543-545, Pisa, Italy, 2004.

G. Attardi, A. Gulli and F. Sebastiani. Automatic Web page categorization by link and context analysis. In Proceedings of THAI'99, European Symposium on Telematics, Hypermedia and Artificial Intelligence, pages 105-119, Varese, Italia, 1999

G. Di Tota, A. Gulli and D. Merletti. SqlWWW: un server Internet per accedere alla base do dati della Soprintendenza ai Beni Ambientali, Architettonici, Artistici e Storici di Pisa, con connessioni di tipoKeep-Alive. In Bollettino VI, n. 1 Scuola Normale Superiore, Pisa, Italy 1998

G. Attardi, A. Gulli and F. Sebastiani THESEUS: categorization by context. In Proceedings of WWW'99, 8th International Conference on the World Wide Web, pages 136-137, Toronto, Canada, 1999

A. Gulli. Jamming.Net: a Server to Balance WWW Load. In Proceedings of WebNet 98 - World Conference on the WWW and Internet & Intranet, Orlando, Florida, USA, 1998

G. Attardi, D. Dato and A.Gulli Web Host Enumeration Through DNS. In Proceedings of Webnet 97 World Conference on the WWW, Internet & Intranet, Toronto, Canada, 1997.

EDUCATION

Università di Pisa, Dipartimento di Informatica

Ph.D. in Computer Science, Dissertation: *Clustering and Ranking for Web Information Retrieval*

Advisor

Professor Paolo Ferragina, Full Professor University of Pisa

External Committee

Professor Andrei Z. Broder, Fellow and VP for Search & Computational Advertising in Yahoo! Research

Professor Tomasz Imielinski, Senior VP, Search Technology at Ask.com

B.S. Degree in Computer Science

B.S. Degree in Engineering

ACADEMIC EXPERIENCE

Università di Pisa, Dipartimento di Informatica

Teaching Assistant Web Information Retrieval Course, 2004-2005

Teaching Assistant System Programming Course, 2005

Organization of Events

Treasurer for 2008 edition of the IEEE International Conference on Data Mining series

PC Member

5th International Conference on Web Search and Data Mining, (wsdm2012), Seattle, 2012

21th International World Wide Web Conference, (www2012), Lyon, 2012

34th European Conference on Information Retrieval. (ECIR2012), Barcelona, 2012

20th International World Wide Web Conference, (www2011), Hyderabad, 2011

18th International World Wide Web Conference, (www2009), Madrid, 2009

31th European Conference on Information Retrieval (ECIR2009), Toulouse, 2009

2nd ACM International Conference on Web Search and Data Mining (WSDM2009), 2009

3rd International ICST Conference on Scalable Information Systems (Infoscale08), Napoli, 2008

4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb08), Beijing, 2008

1st ACM International Conference on Web Search and Data Mining (WSDM2008), Stanford, 2008

30th European Conference on Information Retrieval (ECIR2008), Glasgow, 2008

15th IEEE International Conference on Data Mining series (ICDM 2008), Pisa, 2008

17th International World Wide Web Conference, (www2008), Beijing, 2008

30th Annual International ACM SIGIR Conference, (SIGIR07), Amsterdam, 2007

16th ACM Conference on Information and Knowledge Management (CIKM2007M), Lisbon, 2007

3rd International Workshop on Adversarial Information Retrieval on the Web (AIRWeb07), Banff, 2007

1st Workshop on Text-Mining & Link-Analysis (TextLink-2007), Hyderabad, 2007

29th European Conference on Information Retrieval (ECIR2007), Rome, 2007

16th International World Wide Web Conference (WWW2007), Banff National Park, 2007

Workshop on Link Discovery: Issues, Approaches and Applications (LinkKDD-2006), Philadelphia, 2006

Workshop on Link Discovery: Issues, Approaches and Applications (LinkKDD-2005), Chicago, 2005

Reviewer for Information Processing & Management Journal (IPM)

Reviewer for Applied Mathematics Letters Journal (AML)

FOREIGN
LANGUAGES

Proficient in English and Italian(mother tongue)

TECHNICAL SKILLS

Proficient in C++, STL, Boost, C, Perl. Expert in Python, SQL, Java. Large experience in Search, Machine Learning, and Data mining Loving to innovate

PERSONAL BLOG

<http://codingplayground.blogspot.com/>

random commentary about C++, STL, Boost, Perl, Python, Algorithms and Problem Solving.

OPEN SOURCE
SOFTWARE

Map Browser: a mash-up which geo-localizes your position over a Bing Map and posts it on your Facebook profile with your mood status(HTML5 browser)

Gossip Engine: a search engine which keeps you informed about Gossip events

Picplacer: an engine to geo-localize your facebook pictures

Net Knife: an internet suite for network analysis. More than 10.000 copies downloaded

Jamming.Net a multithread server java for HTTP balancing

MJDBC: a package to optimize the JDBC connections

ClustBoost: a C++ template library for K-means and shingles clustering

Codingplayground A collection of algorithms and data structures in C++ for my playground blog site

- "The best thing about code is, Its simple, easy to implement any where else and all patterns in one place. I have never seen the simplicity of having all patterns in one place. I had been through Modern C++ programming and design patterns from Alexanderscu. You made more easier than that" user comment.
- "Man I like this dudes style. Short and sweet explanations/references of the algorithms and then an implementation in gods own C++. Respek! And thanks for posting!" user comment

REFERENCES

Prof. Andrei Z. Broder, VP of Emerging Search Technology for Yahoo!

Prof. Paolo Ferragina, Full Professor University of Pisa

Prof. Francesco Romani, Full Professor University of Pisa

Dr. Corrado Zoccolo, Google

S. Muthu Muthukrishnan, Google

Antonio Savona, Microsoft

Maxim Gubin, Facebook