

Xiao BAI

Ph.D in Computer Science

Yahoo! Research Barcelona
Avinguda Diagonal 177
08001 Barcelona
☎ +34 (0)693234553
✉ xbai@yahoo-inc.com



Professional Experience

2011-present **Post-doctoral researcher**, *Yahoo! Research Barcelona, Spain.*

Education

- 10/2007 – 12/2010 **Ph.D in Computer Science**, *Institut National des Sciences Appliquées de Rennes, France, mention très honorable.*
Advisor : Dr. Anne-Marie Kermarrec ; Co-advisor : Dr. Marin Bertier
ASAP(As Scalable As Possible) Group, IRISA/INRIA
- 09/2004 – 07/2007 **M.Eng in Computer Science**, *Xi'an Jiaotong University, China, ranked 5th of 56.*
Advisor : Prof. Xiaohong Guan, *IEEE Fellow*
Education Ministry Key Laboratory of Intelligent Network and Network Security of China
- 09/2002 – 07/2004 **Ingénieur**, *Ecole Centrale de Lyon, France, ranked 30th of 320.*
- 09/2000 – 07/2002 **B.Eng in Information Engineering**, *Xi'an Jiaotong University, ranked 2nd of 80.*
Special class for elite students of the grade

Research Experiences

- 09/2011–present **Entity matching.**
Propose an algorithm that co-clusters Web pages and search queries referring to same entities automatically; Propose smoothing technique to tackle with data sparsity; Evaluate the accuracy of matching using large-scale query click logs.
- 02/2011–present **Cache invalidation for real-time indexing, COAST project.**
Work on cache invalidation algorithms to ensure the freshness of query results served from result cache of search engines; Propose a query-driven algorithm that invalidates stale results at query time; Implement the proposed algorithm in an internal search platform and evaluate the performance using realistic workload that preserves the original timing of index updating and query arrival.
- 02/2011–05/2011 **Passive Web crawling through DPI, COAST project.**
Worked on passive Web crawling in geographically distributed search engine; Analyzed network traffic captured through deep package inspection to quantify its impact on traditional Web crawler that discovers new pages through link following; Investigated practical solutions to reduce redundant transmission between DPI nodes and search engine.
- 06/2010–08/2010 **Passive Web crawling through user feedback, COAST project.**
Worked on passive Web crawling in geographically distributed search engine; Analyzed massive user browsing data and click data to quantify the benefits of passive web crawling for increasing the coverage of search engine's document collection and improving the relevance of search results.

- 03/2010 – 06/2010 **On-line personalized query processing in P2P networks**, *GOSSPLE project*.
Designed and implemented a system that enhances the relevance of top-k query processing in P2P networks by building multi-objective overlays using gossip-based protocol ; Proposed a result caching strategy and a query clustering algorithm to improve the efficiency for consecutive and concurrent queries respectively ; Evaluated the performance with real traces from CiteULike and delicious.
- 10/2009 – 03/2010 **On-line personalized top-k processing**, *GOSSPLE project*.
Proposed an algorithm that improves the effectiveness and the efficiency of top-k processing in collaborative tagging systems by an on-the-fly personalization scheme along with a hybrid interest model ; Evaluated the performance using real traces from CiteULike and delicious.
- 09/2008 – 09/2009 **Off-line personalized query processing in P2P networks**, *GOSSPLE project*.
Designed a gossip-based protocol to personalize query processing in social tagging systems that uses implicit user affinities for the personalization and collaboratively collects the query processing information for the efficiency ; Analyzed the optimality theoretically and evaluated the performance using a real trace from delicious.
- 10/2007 – 06/2008 **Network observation**.
Worked over the P2P systems and analyzed the overlay networks using doubling dimension to detect the small-world structures ; Investigated several clustering algorithms to cluster the artists in real trace from Last.fm.
- 03/2005 – 05/2007 **Wireless intrusion detection and prevention**, *Research project in the Education Ministry Key Laboratory of Intelligent Network and Network Security of China*.
Designed and implemented two intrusion detection algorithms and a proactive protection algorithm for three kinds of attacks in 802.11 WLAN.

Scientific Activities

- 09/2011 **Seminar of COAST**, Heidelberg, Germany.
Talk "Active and passive content discovery"
- 09/2010 **Seminar of GOSSPLE**, Rennes, France.
Talk "Narrowing down top-k processing with on-the-fly personalization"
- 08/2010 **Seminar of Yahoo ! Research**, Barcelona, Spain.
Talk "Personalized top-k processing : from centralized to decentralized systems"
- 03/2010 **3rd Winter School on Hot Topics in Distributed Computing**, La Plagne, France.
Talk "Gossiping personalized queries"
- 10/2009 **Seminar of GOSSPLE**, Rennes, France.
Talk "Collaborative personalized top-k processing in P2P systems"
- 07/2009 **Visit in EPFL**, Lausanne, Switzerland.
- 03/2009 **2nd Winter School on Hot Topics in Distributed Computing**, La Plagne, France.
Talk "Decentralized network-aware search"
- 01/2009 **Seminar of GOSSPLE**, Val Thorance, France.
Talk "Decentralized network-aware search"
- 05/2008 **Member of organization committee of 10th francophone conference AlgoTel**, Saint-Malo, France.
- 03/2008 **1st Winter School on Hot Topics in Distributed Computing**, La Plagne, France.

Ph.D Thesis

X. Bai. Personalized top-k processing : from centralized to decentralized systems.

Publications

X. Bai, M. Bertier, R. Guerraoui, and A.-M. Kermarrec. Toward personalized peer-to-peer top-k processing. In *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems (SNS'09)*, pages 1–6.

X. Bai, M. Bertier, R. Guerraoui, A.-M. Kermarrec, and V. Leroy. Gossiping personalized queries. In *Proceedings of the 13th International Conference on Extending Database Technology (EDBT'10)*, pages 87–98.

X. Bai, B. B. Cambazoglu, and F. P. Junqueira. Discovering URLs through user feedback. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM'11)*, pages 77–86.

X. Bai, R. Guerraoui, A.-M. Kermarrec, and V. Leroy. Collaborative personalized top-k processing. To appear in *ACM Transactions on Database Systems*, Volume 36, Issue 4, December 2011.

X. Bai, R. Guerraoui, and A.-M. Kermarrec. Personalizing top-k processing on-the-fly in social tagging networks. *Under submission*.

Grant

2011-2014 **Torres Quevedo Grant**, *Spanish Ministry of Science and Innovation*.

Honors & Awards

2004 – 2005 **Peng-Kang Scholarship & First-Class Scholarship**, *Xi'an Jiaotong University*.

2002 – 2004 **EIFFEL Scholarship**, *French government scholarship for overseas students*.

2000 – 2002 **First-Class Scholarship and Excellent Student Honor**, *Xi'an Jiaotong University*.

2001 – 2002 **Alumnus Yu Bingchang Scholarship**, *Xi'an Jiaotong University*.

Languages

English	TOFEL 620+5	<i>Good command of spoken and written English.</i>
French	DEL F A4	<i>Proficient in spoken and written French, 5 years study in France.</i>
Spanish	Basic	
Chinese	Native language	

Computer Skills

Programming	Proficient in Java and C/C++ on Linux/Windows Proficient in Hadoop MapReduce and Pig Latin Experienced in Matlab
Database	Good knowledge of MySQL Experienced in web crawling and relational database management

Hobbies

Tennis, swimming, skiing and travel