

George Tsatsaronis, Ph.D.

CONTACT INFORMATION

Bioinformatics group
BIOTEC, TU Dresden
Biotechnologisches Zentrum
Tatzberg 47/49
01307 Dresden
Germany



Tel (mobile): +49-15-787 385 954
Tel (work): +49-351-796-5780
Fax (work): +49-351-796-5775
WWW: <http://www.idi.ntnu.no/~ggbt>
E-mail (work): george.tsatsaronis@biotec.tu-dresden.de

PERSONAL DATA

Family Name: Tsatsaronis
First Name: George
Date of birth: October 16, 1979
Place of birth: Athens, Greece
Citizenship: Greek
Marital status: Single

FOREIGN LANGUAGES

Greek

- Fluent - native speaker

English

- Fluent - MICHIGAN Certificate of Proficiency in English (2002), TOEFL, Computer-based exam, score: 273/300 (2002)

German

- A2 level

RESEARCH INTERESTS

Word sense disambiguation, web and text mining, machine learning and artificial intelligence, statistical natural language processing, semantic relatedness and similarity, semantic smoothing kernels, ontologies, semantic graphs, semantic information extraction, digital libraries

CURRENT POSITION

Post-doctoral researcher at the **Bioinformatics Group, BIOTEC**, Dresden Technical University (TU Dresden), Germany.

EDUCATION

Athens University of Economics and Business, Athens, Greece

PhD, Text Mining

Department of Informatics, 2004 - 2009

- PhD Thesis Title: *Word Sense Disambiguation and Text Relatedness based on Word Thesauri.*

- Supervisor: Professor Michalis Vazirgiannis
- Areas of Study: Word sense disambiguation, text mining, measures of semantic relatedness, applications.

University of London, Imperial College, London, U.K.

MSc, Advanced Computing

Department of Computing, 2002 - 2003

- Specialization: Artificial Intelligence and Robotics.
- MSc Thesis Title: *ATLANTIS: A Layered Architecture for Web Content Post Retrieval Analysis*.
- Supervisors: Prof. Yike Guo, Dr. Mustafa Ghanem
- Areas of Study: Search engines, text clustering, word thesauri, web crawling, merging retrieval results, retrieval results visualization.

Athens University of Economics and Business, Athens, Greece

BSc, Informatics

Department of Informatics, 1998 - 2002

- Grade: 8.64/10.0 (< top 1%)
- Areas of Study: Information systems, Databases, Artificial Intelligence, Knowledge Management, Data and Web mining, Web technologies, Networks.

AWARDS/ACHIEVEMENTS/SCHOLARSHIPS/PRIZES

2010

- **Best Paper Award and Best Presentation Award** for the research article *An Experimental Study on Unsupervised Graph-based Word Sense Disambiguation*, presented in the 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2010), March 2010, Iasi, Romania.
- My paper: *A Generalized Vector Space Model for Text Retrieval Based on Semantic Relatedness*, presented at the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009), constitutes one of the basic references in the main *Wikipedia* article: *Generalized Vector Space Model*.

2007

- Scholarship for the 8th EDBT Summer School attendance - Database Technologies for Novel Applications, Bolzano, Italy.

2001

- Prize from *I.K.Y.* for outstanding performance in the 3rd year of studies, Department of Informatics, Athens University of Economics and Business, academic year 2000-2001.
- Scholarship from *GEORGIOS CHALKIPOULOS* Institution, for distinction among the department of informatics students, Athens University of Economics and Business, academic year 2000-2001.

2000

- Scholarship from *GEORGIOS CHALKIOPOULOS* Institution, for distinction among the department of informatics students, Athens University of Economics and Business, academic year 1999-2000.

MEMBERSHIPS

- Member of the Hellenic Artificial Intelligence Society (EETN).
- Member of the American Association for Computational Linguistics (ACL).

PUBLICATIONS

Edited Volumes

V1. *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*, George Paliouras, Constantine Spyropoulos, **George Tsatsaronis** (Eds), Springer Verlag, volume number 6050, released on May 13, 2011.

Book Chapters

BC3. Georgios Paliouras, Constantine D. Spyropoulos, **George Tsatsaronis**, *Bootstrapping Ontology Evolution with Multimedia Information Extraction*, Book Chapter, pp. 1-17, Springer Verlag, vol. 6050, "Knowledge-Driven Multimedia Information Extraction and Ontology Evolution", 2011.

BC2. Magdalini Eirinaki, Dimitris Mavroeidis, **George Tsatsaronis**, Michalis Vazirgiannis, *Introducing Semantics in Web Personalization: The role of Ontologies*, pp. 147-162, Springer, vol. 4289, *Semantics, Web, and Mining*, (Ed.): M. Ackerman, B. Berendt, M. Grobelnik, A. Hotho, D. Mladenic, G. Semeraro, M. Spiliopoulou, G. Stumme, V. Svatek, M. van Someren, 2005.

BC1. Dimitris Mavroeidis. **George Tsatsaronis**, Michalis Vazirgiannis, *Semantic Distances for Sets of Senses and Applications in Word Sense Disambiguation*, pp. 91-105, Knowledge Mining Springer Verlag, Series: Studies in Fuziness and Soft Computing, (Ed.): S. Sirmakessis, 2005.

Journals

J2. Maria Halkidi, Diomidis Spinellis, **George Tsatsaronis**, Michalis Vazirgiannis, *Data Mining in Software Engineering*, Intelligent Data Analysis Journal (IDA), 2011, vol. 15(3), pp. 413-441.

J1. **George Tsatsaronis**, Iraklis Varlamis, Michalis Vazirgiannis, *Text Relatedness based on a Word Thesaurus*, Journal of Artificial Intelligence Research (JAIR), 2010, vol. 37, pp. 1-39.

Conferences

C22. **George Tsatsaronis** *An Experimental Study on Syntactic and Semantic Annotations in Text Retrieval*, CIKM 2011 Workshop on Exploiting Semantic Annotations

(CIKM ESAIR 2011), October 2011, Glasgow, UK.

C21. George Tsatsaronis, Matthias Reimann, Iraklis Varlamis, Orestis Gkorgkas, Kjetil Nørnvåg *Efficient Community Detection Using Power Graph Analysis*, CIKM 2011 Workshop on Large-Scale and Distributed Systems for Information Retrieval (CIKM LSDS-IR 2011), October 2011, Glasgow, UK.

C20. Robert Neumayer, **George Tsatsaronis**, Kjetil Nørnvåg *TRUMIT: A Tool to Support Large-Scale Mining of Text Association Rules*, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2011), September 2011, Athens, Greece.

C19. George Tsatsaronis, Iraklis Varlamis, Sunna Torge, Matthias Reimann, Kjetil Nørnvåg, Michael Schroeder, Matthias Zschunke *How to Become a Group Leader? or Modelling Author Types based on Graph Mining*, International Conference on Theory and Practice of Digital Libraries (TPDL 2011), September 2011, Berlin, Germany.

C18. George Tsatsaronis, Natalia Macari, Sunna Torge, Heiko Dietze, Michael Schroeder, *A Maximum Entropy Approach for Accurate Document Annotation in the Biomedical Domain*, ISMB/ECCB BioOntologies Special Interest Group 2011 (BioOntologies 2011), July 2011, Vienna, Austria.

C17. Jamal Abdul Nasir, Asim Karim, **George Tsatsaronis**, Iraklis Varlamis, *A Knowledge-based Semantic Kernel for Text Classification*, 18th International Symposium on String Processing and Information Retrieval (SPIRE 2011), October 2011, Pisa, Italy.

C16. Iraklis Varlamis, **George Tsatsaronis**, *Visualizing Bibliographic Databases as Graphs and Mining Potential Research Synergies*, 2011 International Conference on Social Networks Analysis and Mining (ASONAM 2011), July 2011, Taiwan.

C15. George Tsatsaronis, Iraklis Varlamis, Kjetil Nørnvåg, *SemanticRank: Ranking Keywords and Sentences Using Semantic Graphs*, 23rd International Conference on Computational Linguistics (COLING 2010), August 2010, Beijing, China.

C14. Katerina Papantoniou, **George Tsatsaronis**, George Paliouras, *KDTA: Automated Text Annotation with Knowledge-Driven Text Analysis*, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2010), September 2010, Barcelona, Spain.

C13. George Tsatsaronis, Iraklis Varlamis, Kjetil Nørnvåg, *An Experimental Study on Unsupervised Graph-based Word Sense Disambiguation*, **Best Paper Award and Best Presentation Award**, 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2010), March 2010, Iasi, Romania.

C12. Elias Zavitsanos, **George Tsatsaronis**, Iraklis Varlamis, George Paliouras, *Scalable Semantic Annotation of Text Using Lexical and Web Resources*, 6th Hellenic Conference on Artificial Intelligence (SETN 2010), May 2010, Athens, Greece.

C11. George Tsatsaronis, Iraklis Varlamis, Andreas Giannakopoulos, Nikolaos Kanellopoulos, *Identifying Free Text Plagiarism Based on Semantic Similarity*, 4th International Conference on Plagiarism Detection, July 2010, Newcastle, UK.

C10. George Tsatsaronis, Iraklis Varlamis, Sofia Stamou, Nørnvåg, Michalis Vazir-

giannis, *Semantic Relatedness Hits Bibliographical Data*, 11th ACM International Workshop on Web Information and Data Management (WIDM 2009), pp. 87-90, November 2nd 2009, HongKong, China.

C9. George Tsatsaronis, Iraklis Varlamis, Michalis Vazirgiannis, Kjetil Nørvåg, *Omiotis: A Thesaurus-based Measure of Text Relatedness*, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2009), pp. 742-745, 7-11 September 2009, Bled, Slovenia.

C8. George Tsatsaronis, Vicky Panagiotopoulou, *A Generalized Vector Space Model for Text Retrieval Based on Semantic Relatedness*, 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009), pp. 70-78, 30 March -3 April 2009, Athens, Greece.

C7. George Tsatsaronis, Maria Halkidi, Emmanouel Giakoumakis, *Quality Classifiers for Open Source Software Repositories*, AIAI 2009 2nd Artificial Intelligence Techniques in Software Engineering Workshop, 22-23 April 2009, Thessaloniki, Greece.

C6. George Tsatsaronis, Iraklis Varlamis, Michalis Vazirgiannis, *Word Sense Disambiguation with Semantic Networks*, 11th International Conference on Text, Speech and Dialogue (TSD 2008), pp. 219-226, 8-12 September 2008, Brno, Czech Republic.

C5. George Tsatsaronis, Michalis Vazirgiannis, Ion Androutopoulos, *Word Sense Disambiguation with Spreading Activation Networks Generated from Thesauri*, 20th International Joint Conference in Artificial Intelligence, (IJCAI 2007), AAAI Press, pp. 1725-1730, 6-12 January 2007, Hyderabad, India.

C4. Nikos Zotos, Paraskevi Tzekou, **George Tsatsaronis**, Lefteris Kozanidis, Sofia Stamou, Iraklis Varlamis, *To Click or not to Click? The Role Of Contextualized and User-Centric Web Snippets*, SIGIR 2007 Workshop on Focused Retrieval, July 27, 2007, Amsterdam, The Netherlands.

C3. Dimitris Mavroeidis, **George Tsatsaronis**, Michalis Vazirgiannis, Martin Theobald, Gerhard Weikum, *Word Sense Disambiguation for Exploiting Hierarchical Thesauri in Text Classification*, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2005), pp. 181-192, 3-7 October 2005, Porto, Portugal.

C2. George Tsatsaronis, Reetta Pitkanen, Michalis Vazirgiannis, *Clustering for Ontology Evolution*, 29th Annual Conference of the German Classification Society (GfKI 2005), p. 167, 9-11 March, Magdeburg, Germany.

C1. Dimitris Mavroeidis, **George Tsatsaronis**, Michalis Vazirgiannis, *Semantic Distances for Sets of Senses and Applications in Word Sense Disambiguation*, NEMIS, 3rd International Workshop in Text Mining and its Applications, 25 October 2004, Athens, Greece.

Technical Reports

T2. Michalis Vazirgiannis, Maria Halkidi, **George Tsatsaronis**, Evripides Vrachnos, *Recent Advances on Pattern Representation and Management*, PANDA Project Technical Report TR-2003-04, December 2003.

T1. Michalis Vazirgiannis, Maria Halkidi, **George Tsatsaronis**, Evaripides Vrachnos, *A Survey on Pattern Application Domains and Pattern Management Approaches*, PANDA Project Technical Report TR-2003-01, February 2003.

ACADEMIC EXPERIENCE

Post-Doc Fellow, March 2010 to present
Department of Computing, Technical University of Dresden, Biotechnology Center (BIOTEC)

- Areas of Research: Text Mining, Natural Language Processing, Artificial Intelligence, Machine Learning, Bioinformatics

Post-Doc Fellow, May 2009 to March 2010
Department of Computer and Information Science, Norwegian University of Science and Technology

- Areas of Research: Distributed Text Mining, Natural Language Processing, Machine Learning, Artificial Intelligence.
- Participation in the COMIDOR Norwegian project, which aims at design and development of distributed text mining techniques.

Research Associate, 20/11/2008 - Present

Research associate of the National Center of Scientific Research, Institute of Informatics and Telecommunications.

Research Associate, 1/5/2002 - 1/2/2010

Research associate of the DB-NET research team in Data and Web mining, Department of Informatics, Athens University of Economics and Business.

Participation in Research Projects

Participation in 7 research projects funded by the E.U.:

- 2001-2004 IST/FET - DB/GLOBE (IST-2001-32645)
- 2002-2004 IST/FET - PANDA (IST-2001-33058)
- 2002-2004 IST/H-TECHSIGHT (IST-2001-33174)
- 2001-2003 IST/I-KnowUMine (IST-2000-31077)
- 2006-2008 SQO-OSS (IST-5-033331)
- 2006-2009 BOEMIE (FP6-027538)
- 2008 - 2011 CASAM (FP7-217061)

Citations

- 81 (up to 30/04/2011 - excluding self citations)
- h-index: 6 (up to 30/04/2011 - excluding self citations)

Program Committee/Organizing Committee

- Organizing Committee Member of the 2011 ECML/PKDD Workshop on Mining Complex Entities from Network and Biomedical Data (MIND 2011), held in conjunction with the 2011 ECML/PKDD Conference, September 5-9, 2011, Athens, Greece.
- Program Committee Member of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011).
- Program Committee Member of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011).
- Program Committee Member of the 12th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2011).
- Organizing Committee Member of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009).

Supervision

- Co-supervisor of PhD student Orestis Gkorgas (three member committee), as of July 2010, Department of Computer and Information Science, Norwegian University of Science and Technology. Estimated PhD award date: end of 2014.
- Co-supervisor of Natalia Macari's MSc thesis, Department of Computing, Technical University of Dresden, June 2010 - December 2010.
- Supervisor of Vicky Panagiotopoulous's MSc thesis, Department of Informatics, Athens University of Economics and Business, December 2008 - May 2009.
- Supervisor of 4 BSc thesis, Department of Informatics, Athens University of Economics and Business, during the period 2005 - 2009.

External Reviewer

- Journals:
 - IEEE Intelligent Systems
 - IEEE Transactions on Neural Networks
 - BMC Bioinformatics
 - Nucleid Acids Research (Oxford Journals)
 - IEEE Transactions on Systems, Man and Cybernetics
 - Engineering Applications of Artificial Intelligence (Elsevier)

- IOS Intelligent Data Analysis Journal
- Conferences:
 - International Conference on Theory and Practice of Digital Libraries (TPDL 2011)
 - 19th Annual International Conference on Intelligent Systems for Molecular Biology and 10th European Conference on Computational Biology (ISMB-ECCB 2011)
 - 14th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2010)
 - 18th Int. World Wide Web Conference (WWW 2009)
 - 6th European Semantic Web Conference (ESWC 2009)
 - 13th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2009)
 - 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009)
 - IEEE International Conference on Data Mining 2008 (ICDM 2008)
 - 12th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2008)
 - ACM SIGMOD Conference 2008
 - 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2007)

Teaching Experience

- Department of Computing, Technical University of Dresden
 - 1/2 Master Course, in collaboration with Prof. Dr. Michael Schroeder. Course Title: *Semantic Search*, Spring Semester 2011.
 - 1/2 Master Course, in collaboration with Prof. Dr. Michael Schroeder. Course Title: *Integrated Logic Systems*, Spring Semester 2010.
- Department of Computer and Information Science, Norwegian University of Science and Technology
 - T.A. - Delivery of two weeks lectures on measures of evaluation in information retrieval, under the framework of the respective undergraduate course of Information Retrieval, October 2009.
- Department of Informatics, Athens University of Economics and Business
 - T.A. - Lab in Data and Web Mining technologies, 18 hours, in the framework of the undergraduate course Data and Web Mining, September 2004 - January 2005.
 - T.A. - Tutorial in Database design, 18 hours, in the framework of the undergraduate course Database Design, September 2004 - January 2005.
 - 10 hour seminar *Introduction in Database design*, in the framework of the MSc program study MSc in Real Estate, Department of Statistics, A.U.E.B., September 2004.
 - T.A. - Tutorial in Database design, 18 hours, in the framework of the undergraduate course Database Design, September 2005 - January 2006.

- T.A. - Tutorial in Artificial Intelligence, 18 hours, in the framework of the undergraduate course Artificial Intelligence, September 2005 - January 2006.
- T.A. - Lab in Database and Web applications, 18 hours, in the framework of the MSc course Databases, MSc in Information Systems, A.U.E.B., September 2005 - January 2006.
- T.A. - Tutorial in Artificial Intelligence, 10 hours, in the framework of the undergraduate course Artificial Intelligence, September 2006 - January 2007.
- T.A. - Lab in C++, 10 hours, in the framework of the undergraduate course Computer Programming with C++, September 2006 - January 2007.
- T.A. - Tutorial in Artificial Intelligence, 24 hours, in the framework of the undergraduate course Artificial Intelligence, September 2007 - January 2008.
- T.A. - Lab in C++, 24 hours, in the framework of the undergraduate course Computer Programming with C++, September 2007 - January 2008.

Participated Conferences

- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2010), September 20-24, 2010, Barcelona, Spain
- 23rd International Conference on Computational Linguistics (Coling 2010), August 23-27, 2010, Beijing, China.
- 6th Hellenic Conference on Artificial Intelligence (SETN), May 2010, Athens, Greece.
- 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), March 2010, Iasi, Romania.
- 18th ACM Conference on Information and Knowledge Management (CIKM), November 2-6, 2009, Hong Kong, China.
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), September 7-11, 2009, Bled, Slovenia.
- 12th Conference of the European Chapter of the Association for Computational Linguistics. March 30 - April 3, 2009, Athens, Greece.
- 11th International Conference on Text, Speech and Dialogue, September 8-12, 2008, Brno, Czech Republic.
- 30th Annual International ACM SIGIR Conference (SIGIR 2007), July 23-27, 2007, Amsterdam, The Netherlands.
- 20th International Joint Conference on Artificial Intelligence (IJCAI 2007), January 6-12, 2007, Hyderabad, India.
- 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2005), 3-8 October 2005, Porto, Portugal.
- 6th International Venture Capital Forum, Hotel Grande Bretagne, 14-15 June 2005, Athens, Greece.
- 29th Annual Conference of the German Classification Society (GfKl 2005), 9-11 March 2005, Magdeburg, Germany.

PROFESSIONAL EXPERIENCE

- March 2010 - present, IT Consultant at *Transinsight GmbH*, responsible for new semantic search technologies.
- Participation in the PRAKSE A national project, concerning the creation of a spin-off data mining software company, providing respective products and services.
- February 2004 - June 2009, freelancer with profession objects: 1) Implementation services for software systems and tools, 2) Research services in computer science. In the framework of this professional activity, the following have been compiled:
 - Participation in a PAVET national project Hedge Funds - Portfolio information management system, in co-operation with Can Communications (<http://www.cancom.gr/>), April - June 2005.
 - Participation in design and implementation of Vasdekis Editions Web page, July 2004.
 - Design and implementation of an on-line livescore soccer results information system, March 2004.
 - Design and implementation of real estate adds management system, in co-operation with Department of Statistics, A.U.E.B., February 2004.

SEMINARS/SUMMER SCHOOLS

- 8th EDBT Summer School - Database Technologies for Novel Applications, September 3-7, 2007, Bozen-Bolzano, Italy.
- Microsoft Certificate of Achievement on *Microsoft Visual Studio .Net* (2002).

TECHNICAL SKILLS

- Operating Systems: MS Windows 98/98SE/M/2000/NT/XP/VISTA/2003/2008/7, Ubuntu, Unix, Linux Red Hat
- Programming Languages: Java, C, C++, JDBC, TCL/Tk, PERL, JSP, PHP, ASP, Java Script, VB, Assembly, UNIX shell, SQL, HTML, XML, XHTML, WML, HT- DML, Pascal, Fortran, Cobol.
- DBMSs: MS SQL SERVER 2000/2005/2008, Oracle 9i/10g, MS ACCESS, POSTGRESS SQL, MySQL.
- Packages: MS OFFICE 97/00/03/07, Visual Studio 6, Electronic Workbench, SPSS, Magnum Opus, Cognos, U-miner, Nokia Wap Toolkit, Ericsson MPS, Ericsson WAP IDE 3.1.1, KENSINGTON DISCOVERY EDITION (KDE), SPSS Clementine, Weka, SVM Light.

OTHER SKILLS

- Excellent communication skills.
- Excellent presentation skills.
- Ability and experience to write research proposals, attract and expand collaborations, and attract research funding.
- Managing skills of large research projects.

CAREER TARGETS

Academic Career in Computer Science, Technology Consultant