ALI CAKMAK

Department of Computer Engineering
Istanbul Technical University
Ayazaga Campus
34467 Maslak, İstanbul

ali.cakmak@itu.edu.tr
http://web.itu.edu.tr/alicakmak

EDUCATION

Case Western Reserve University, Cleveland, Ohio
Doctor of Philosophy, Computer Science
<div align="right">Aug. 2003
-Sept. 2008</div>

Research interests:

Bioinformatics, Applied Machine Learning, Data Science, Data Mining, Medical Informatics, Data Management, Databases, Query Optimization

Bilkent University, Ankara, Turkey
Bachelor of Science, Computer Engineering
<div align="right">Sept. 1999
-May 2003</div>

PROFESSIONAL EXPERIENCE

Istanbul Technical University, Istanbul, Turkey
Associate Professor, Vice Chair
Department of Computer Engineering
<div align="right">Sept. 2020,
- Present</div>

Health Institutes of Turkey, Istanbul, Turkey
Consulting Researcher
Institute of Artificial Intelligence and Health Data Research
<div align="right">April 2022,
- Present</div>

Istanbul Sehir University, Istanbul, Turkey
Associate Professor
Department of Computer Engineering and Science
<div align="right">Sept. 2019,
- Sept. 2020</div>

Istanbul Sehir University, Istanbul, Turkey
Assistant Professor
Department of Computer Engineering and Science
<div align="right">Aug. 2013
- June 2019</div>

Oracle, Inc., Redwood Shores, California, USA
Senior Member of Technical Staff
Query Optimizer Team, Database Technologies
<div align="right">Nov. 2009
- July 2013</div>

Case Western Reserve University, Cleveland, Ohio, USA
Instructor & Postdoctoral Research Associate
Department of Electrical Engineering and Computer Science
<div align="right">Oct. 2008
- Oct. 2009</div>

Case Western Reserve University, Cleveland, Ohio, USA
Research Assistant, Bioinformatics and Database Research Group
Department of Electrical Engineering and Computer Science
<div align="right">Fall 2003
- Fall 2008</div>

RESEARCH EXPERIENCE

Istanbul Technical University & Istanbul Sehir University, Istanbul, Turkey
<div align="right">Aug. 2013
- Present</div>

- *Bioinformatics:* Developing system biology algorithms for the automated interpretation of metabolomics data, and later link it to possible diagnosis alternatives for personalized medicine -- Scalable taxonomy classification with hierarchical supervised models -- Drug target prediction and drug repurposing with multi-omics data.

- *Health Informatics*: Computational tools for the diagnosis of non-alcoholic fatty liver disease based on blood tests – Estimating the tendency to colorectal cancer using SNP Profiles – Developing a risk monitor for patient readmissions at hospitals.

- *Medical Informatics*: Decision support models for clinicians to make the optimal discharge time decision based to minimize hospital readmissions for chronic diseases.

- *Database Research*: Estimating the selectivity of fuzzy string queries

- *Sports Analytics*: Analyze game data to reveal non-obvious team- and player-level patterns to guide instructors and players for performance improvement, and designing adaptive game tactics.

Oracle, Inc., Redwood Shores, California, USA
Senior Member of Technical Staff
<div align="right">Nov. 2009 - July 2013</div>

- *Estimating Database Statistics On-the-fly:* Research on learning statistics patterns, and adjusting previously collected stats dynamically during query compile-time according to the trends discovered based on historical data.

- *New histogram types:* Introduced two new histograms (i.e., top-frequency and hybrid) which provide more accurate predicate selectivity estimates during query optimization.

- *Concurrent statistics gathering*: Developed a framework that allows gathering statistics for multiple objects concurrently. Improved statistic s gathering time up to 10x on the Oracle database machine.

Case Western Reserve University, Cleveland, Ohio, USA
Postdoctoral Research Associate
<div align="right">Oct. 2008 - Oct. 2009</div>

- *Systems Biology Research*: Participated in the design and development of database-enabled framework and tools to facilitate effective and efficient model development for multiscale mechanistic models of biological systems. Contribute to the PathCase-SB system's interfaces involving Systems Biology Models-based querying, visualization, simulation, and model building.

- *Metabolomics Research*: Design and query processing of a query language for mammalian metabolic networks, development of graph-based metabolomics data analysis techniques.

Case Western Reserve University, Cleveland, Ohio, USA
Research Assistant, Database and Bioinformatics Research Group
<div align="right">Fall 2003 - Fall 2008</div>

- *Bioinformatics Research:* Functionality annotation analysis of metabolic pathways with applications on pathway inference and categorization, metabolomics data analysis.

- *Data Mining Research:* Mining for frequent graph structures in taxonomy super-imposed graph databases

- *Information Extraction Research:* Extracting gene annotations from biomedical literature using textual annotation patterns and semantic pattern matching techniques. Design of a next-generation novel search engine that collects and mines for information about individuals on the WWW (PopulusLog)

- *Information Retrieval Research:* Investigating the context-based search paradigm, evaluation of publication similarity measures, exploring different ranking functions.

- *Databases Research:* Data Integration with particular focus on biological web data sources, data and visualization models for pathways, pathways databases

system with advanced querying and visualization tools.

- *Metabolomics Research:* Metabolic-network projected automated interpretation of concentration changes for a large number of metabolites. Developed models for hypothesis generation, consistency checking, and ranking of alternative hypotheses based on observed metabolite changes.

TEACHING EXPERIENCE

Istanbul Technical University, Istanbul, Turkey

Instructor – BLGE 565: Bioinformatics Algorithms
Graduate Level (Elective) (ECTS: 7.5)

Instructor – BLGE 564: Advanced Database Systems
Graduate Level (Elective) (ECTS: 7.5)

Instructor – BLG348E: Introduction to Bioinformatics
Junior Level (Elective) (ECTS: 5)

Instructor – BLG317E: Database Systems
Junior Level (Mandatory) (ECTS: 4.5)

Instructor – BLGE 101: Intro. to Information Systems and Comp. Eng.
Freshman Level (Mandatory) (ECTS: 5)

Instructor – BLGE 102: Intro. to Programming
Freshman Level (Mandatory) (ECTS: 8)

Istanbul Sehir University, Istanbul, Turkey

Instructor – ECE 527: Bioinformatics
Graduate Level (Elective) (ECTS: 7.5)

Instructor – CS 350: Database Systems
Junior Level (Mandatory) (ECTS: 5)

Instructor – ENGR 101: Introduction to Programming
Freshman Level (Mandatory) (ECTS: 5)

Instructor – ENGR 102: Programming Practice
Freshman Level (Mandatory) (ECTS: 5)

Instructor – ECE 562: Networks Modeling
Graduate Level (Elective) (ECTS: 7.5)

Case Western Reserve University, Cleveland, Ohio

Instructor – EECS 233: Data Structures
Sophomore Level (Mandatory) (ECTS: 5)

DEVELOPMENT EXPERIENCE

Oracle, Inc., Redwood Shores, California, USA
Senior Member of Technical Staff

Nov. 2009
- July 2013

- Member of the Query Optimizer Group in the Oracle Database product. Focused on the optimizer statistics management area within the group. Also contributed to the maintenance and enhancement of all aspects of the query optimizer within the RDBMS kernel.

Case Western Reserve University, Cleveland, Ohio, USA
Database and Bioinformatics Research Group

Fall 2003
- Fall 2009

- *PathCase Metabolic Pathways Database System:* A bioinformatics tool to store, query, analyze, and visualize biological pathways. Took part in the design, implementation, and coordination phases.

- *CaseExplorer, The Case Digital Library Project:* An umbrella project housing online tools to effectively query and rank scientific papers. Developed Pubmed Abstract Full-Text Search Prototype running on 7,000,000 paper abstracts, and took part in the development of similarity and ranking measures for other tools.

- *PopulusLog People Information Database:* A web data mining research project and the associated knowledgebase to collect publicly available online information about people. Provides tools for crawling web pages, information extraction, querying, and visualization of automatically constructed social networks.

PUBLICATIONS

Patents:

Pt1. Chakkappen, S. P., Zait, M., Lee, A. W., & Cakmak, A. (2022). U.S. Patent No. 11,403,295. Washington, DC: U.S. Patent and Trademark Office.

Pt2. Delibas, E., Uzun, A., Inan, M. F., Guzey, O., & Cakmak, A. (2019). *U.S. Patent Application No. 16/331,744.*

Pt3. Cakmak, A., Celik, M.H. (2018). A System for Diagnosing Diseases. *International Patent App. No. PCT/TR2019/051101.*

Pt4. Cakmak, A., Aytimur, M. (2018). LIKE Selectivity Estimation. *International Patent App. No. PCT/TR2018/050366.*

Pt5. Chakkappen, S. P., Zait, M., Lee, A. W., & Cakmak, A. (2016). *U.S. Patent No. 9,471,631.* Washington, DC: U.S. Patent and Trademark Office.

Refereed/Invited Journal Publications:

J1. Cakmak, A., Ibrahimzada A. R., Arıkan, S., Ayaz, H., Demirkol, Ş., Sönmez, D., Hakan, M. T., Turan, S. S., Horozoğlu, C., Küçükhüseyin, Ö., Kiran, B., Zeybek, Ş. Ü., Baysan, M., Yaylim, I. (2022). Predicting the Predisposition to Colorectal Cancer based on SNP Profiles of Immune Phenotypes using Supervised Learning Models. Medical & Biological Engineering & Computing (SCI) (in press).

J2. Cakmak, A., & Celik, M. (2021). Personalized Metabolic Analysis of Diseases. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 18(3), 1014-1025. (SCI).

J3. Aytimur, M., & Cakmak, A. (2021). Using positional sequence patterns to estimate the selectivity of SQL LIKE queries. Expert Systems with Applications, 165, 113762. (SCI).

J4. Sohsah, G., Ibrahimzada, A. R., Ayaz, H., & Cakmak, A. (2020). Scalable Classification of Organisms into a Taxonomy Using Hierarchical Supervised Learners. Journal of Bioinformatics and Computational Biology (5):2050026 (SCI-E).

J5. Delibas, E., Uzun, A., Inan, M. F., Guzey, O., & Cakmak, A. (2019). Interactive exploratory soccer data analytics. INFOR: Information Systems and Operational Research, 57(2), 141-164. (SCI-E).

J6. Polat, N., Cakmak, A., & Turan, R. (2020). Exploring the power of supervised learning methods for company name disambiguation in microblog posts. Turkish Journal of Electrical Engineering & Computer Sciences, 28(5), 2400-2415.

J7. Cakmak, A., Uzun, A., & Delibas, E. (2018). Computational modeling of pass effectiveness in soccer. Advances in Complex Systems, 21(03n04), 1850010. (SCI-E).

J8. Aytimur, M., & Çakmak, A. (2018). Estimating the selectivity of LIKE queries using pattern-based histograms. Turkish Journal of Electrical Engineering & Computer

Sciences, 26(6), 3319-3334. (SCI-E).

J9. Cakmak, A. (2017). Predicting student success in courses via collaborative filtering. International Journal of Intelligent Systems and Applications in Engineering, 5(1), 10-17.

J10. Cicek, A.E., Qi, X., Cakmak, A., Johnson, S.R., Han, X., Alshalwi, S., Ozsoyoglu, Z.M. and Ozsoyoglu, G. (2014). An online system for metabolic network analysis. Database, 2014. (SCI-E)

J11. Qi, X., Coskun, S.A., Cakmak, A., Cheng, E., Cicek, E.A., Das, M., Yang, L., Jadeja, R., Syed, S., Dash, R.K. and Lai, N. (2012). PathCase-SB: database-enabled tools for regulatory metabolic networks. ACM SIGBioinformatics Record, 2(3), pp.19-25.

J12. Cakmak, A., Qi, X., Cicek, A. E., Bederman, I., Henderson, L., Drumm, M., & Ozsoyoglu, G. (2012). A new metabolomics analysis technique: steady-state metabolic network dynamics analysis. Journal of bioinformatics and computational biology, 10(01), 1240003. (SCI-E).

J13. Coskun, S.A., Qi, X., Cakmak, A., Cheng, E., Cicek, A.E., Yang, L., Jadeja, R., Dash, R.K., Lai, N., Ozsoyoglu, G. and Ozsoyoglu, Z.M. (2012). PathCase-SB: integrating data sources and providing tools for systems biology research. BMC systems biology, 6(1), p.67. (SCI).

J14. Cakmak, A., Qi, X., Coskun, S.A., Das, M., Cheng, E., Cicek, A.E., Lai, N., Ozsoyoglu, G. and Ozsoyoglu, Z.M. (2011). PathCase-SB architecture and database design. BMC systems biology, 5(1), p.188. (SCI).

J15. Cakmak, A., Ozsoyoglu, G., & Hanson, R. W. (2010). Querying metabolism under different physiological constraints. Journal of bioinformatics and computational biology, 8(02), 247-293. (SCI-E).

J16. Ratprasartporn, N., Po, J., Cakmak, A., Bani-Ahmad, S., & Ozsoyoglu, G. (2009). Context-based literature digital collection search. The VLDB Journal, 18(1), 277-301. (SCI).

J17. Elliott, B., Kirac, M., Cakmak, A., Yavas, G., Mayes, S., Cheng, E., Wang, Y., Gupta, C., Ozsoyoglu, G. and Meral Ozsoyoglu, Z. (2008). PathCase: pathways database system. Bioinformatics, 24(21), pp.2526-2533. (SCI).

J18. Cakmak, A., & Ozsoyoglu, G. (2008). Discovering gene annotations in biomedical text databases. BMC bioinformatics, 9(1), 143. (SCI-E).

J19. Cakmak, A., & Ozsoyoglu, G. (2007). Mining biological networks for unknown pathways. Bioinformatics, 23(20), 2775-2783. (SCI).

J20. Bani-Ahmad, S., Cakmak, A., Özsoyoglu, G., & Al-Hamdani, A. (2005). Evaluating Publication Similarity Measures. *IEEE Data Eng. Bull.*, *28*(4), 21-28.

Refereed Conference/Workshop Publications:

C1. Cakmak, A., Ibrahimzada A. R., Arıkan, S., Ayaz, H., Demirkol, Ş., Sönmez, D., Hakan, M. T., Turan, S. S., Horozoğlu, C., Küçükhüseyin, Ö., Kiran, B., Zeybek, Ş. Ü., Baysan, M., Yaylim, I. (2019). Predicting the Predisposition to Colorectal Cancer. VII. International Molecular Medicine Congress.

C2. Cakmak, A., Celik, MH. (2018). Personalized metabolic analysis of diseases. The International Symposium on Health Informatics and Bioinformatics (HIBIT). October 25-27, 2018.

C3. Davletov, F., Aydin, A. S., & Cakmak, A. (2014, November). High impact academic paper prediction using temporal and topological features. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (pp.

491-498).

C4. Bulut, A., Arıcı, T., Guzey, O., Arslan, B., Cakmak, A. (2014, October). Training the Data Scientists of the Future". Big Data and Analytics EdCon, Las Vegas, Nevada, USA.

C5. Cicek, A. E., Qi, X., Cakmak, A., Johnson, S. R., Han, X., Alshalwi, S., & Ozsoyoglu, G. (2013, September). PathCase-MAW: An online metabolic network analysis workbench. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics* (pp. 219-228).

C6. Cakmak, A., Qi, X., Cicek, A. E., & Özsoyoğlu, G. (2011, August). Computational interpretation of metabolomics measurements: Steady-state metabolic network dynamics analysis. In *Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine* (pp. 387-392).

C7. Elliott, B., Mayes, S., Cakmak, A., Ozsoyoglu, G., & Ozsoyoglu, Z. M. (2010, March). Advanced querying interface for biochemical network databases. In Proceedings of the 2010 ACM Symposium on Applied Computing (pp. 1526-1533).

C8. Cakmak, A., Dsouza, A., Hanson, R. W., & Ozsoyoglu, M. (2010, August). Analyzing metabolite measurements for automated prediction of underlying biological mechanisms. In *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology* (pp. 137-146).

C9. Cheng, E., Cakmak, A., & Ozsoyoglu, Z. M. (2010, August). Efficient query evaluation for DAG-shaped hierarchies. In Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology (pp. 454-456).

C10. Cakmak, A., Ozsoyoglu, G., & Hanson, R. W. (2009, August). Managing and Querying Mammalian Metabolic Networks: A Metabolism Query Language and Its Query Processing. In *Proceedings of 8th International Conference on Computational Systems Bioinformatics. Stanford, CA*.

C11. Cakmak, A., Dsouza, A., Hanson, R., Ozsoyoglu, G., & Ozsoyoglu, M. (2009, September). A web-based data source for metabolomics. In *2009 24th International Symposium on Computer and Information Sciences* (pp. 454-459). IEEE.

C12. Cakmak, A., Kirac, M., & Ozsoyoglu, G. (2009, September). PopulusLog: People information database. In *2009 24th International Symposium on Computer and Information Sciences* (pp. 165-170). IEEE.

C13. Cakmak, A., & Ozsoyoglu, G. (2008, March). Taxonomy-superimposed graph mining. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology* (pp. 217-228).

C14. Cakmak, A., Kirac, M., Reynolds, M. R., Ozsoyoglu, Z. M., & Ozsoyoglu, G. (2007, July). Gene ontology-based annotation analysis and categorization of metabolic pathways. In *19th International Conference on Scientific and Statistical Database Management (SSDBM 2007)* (pp. 33-33). IEEE.

C15. Ratprasartporn, N., Bani-Ahmad, S., Cakmak, A., Po, J., & Ozsoyoglu, G. (2007, April). Evaluating different ranking functions for context-based literature search. In *2007 IEEE 23rd International Conference on Data Engineering Workshop* (pp. 261-268). IEEE.

C16. Cakmak, A., & Ozsoyoglu, G. (2007). Annotating genes using textual patterns. In Biocomputing 2007 (pp. 221-232).

C17. Kirac, M., Cakmak, A., & Ozsoyoglu, G. (2006, July). Task-Oriented Integrated Use of Biological Web Data Sources. In *18th International Conference on Scientific and Statistical Database Management (SSDBM'06)* (pp. 81-90). IEEE.

C18. Ratprasartporn, N., Cakmak, A., & Ozsoyoglu, G. (2006, July). On data and visualization models for signaling pathways. In *18th International Conference on Scientific and*

Statistical Database Management (SSDBM'06) (pp. 133-142). IEEE.

C19. Bani-Ahmad, S., Cakmak, A., Al-Hamdani, A., & Ozsoyoglu, G. (2005, December). Evaluating score and publication similarity functions in digital libraries. In *International Conference on Asian Digital Libraries* (pp. 483-485). Springer, Berlin, Heidelberg.

Other Publications:

O1. Belknap, P., Cakmak, A., Chakkappen, S., Chan, I., Chatterjee, D., Das, D., ... & Lee, A. (2013). Oracle Database SQL Tuning Guide, 12c Release 1 (12.1) E15858-15.

RESEARCH GRANTS

Awarded:

- TUSEB TA-01, "Tools and Algorithms for Multi-omics Supported Personalized Treatment Recommendation, Drug Repositioning, and Drug Target Discovery". The Health Institutes of Turkey (Role: PI, Budget: ~349,000 TL, Duration: 36 Months) — **2022 - 2024**

- TUSEB TA-01, "A Flexible and Easy-to-use Pipeline for Next Generation Sequencing Analysis of Cancer Samples", The Health Institutes of Turkey (Role: Co-PI, Budget: ~348,000 TL, Duration: 36 Months) — **2020 - 2023**

- TUBITAK ARDEB 1001, "New Methods and Algorithms to Estimate the Selectivity of SQL LIKE Queries", The Scientific and Technological Research Council of Turkey (Role: PI, Budget: ~121,000 TL, Duration: 21 Months) — **2017 - 2019**

- TUBITAK ARDEB 3501 (CAREER Grant), "Algorithms and Tools for Computational Modeling of Metabolomics Data", The Scientific and Technological Research Council of Turkey (Role: PI, Budget: ~223,000 TL, Duration: 27 Months) — **2015 - 2017**

- TUBITAK BIDEB 2232 (Returning Researcher Grant), "Data Mining Techniques for Fast Query Optimization in Relational Database Systems", The Scientific and Technological Research Council of Turkey (Role: PI, Budget: ~95,000 TL, Duration: 24 Months) — **2014 -2015**

Under Review/Development:

- "Readmission Risk-based Discharge and Planned Admission Decision Support System", EU Horizon (Role: PI, Budget: ~1,740,000 Euro, Duration: 60 Months) (To be submitted in January 2022)

ACTIVITIES

- Program Committee Member, The International Symposium on Health Informatics and Bioinformatics (HIBIT 2021) — **2021**

- Program Committee Member, The International Symposium on Health Informatics and Bioinformatics (HIBIT 2020) — **2020**

- Program Committee Member, International Conference on Internet of Things, — **2018**

Big Data, and Security (IoTBDS 2018)

- Program Committee Member, International Conference on Information, Process, and Knowledge Management (eKNOW 2018) **2018**

- Program Committee Member, International Conference on Information, Process, and Knowledge Management (eKNOW 2017) **2017**

- Program Committee Member, International Conference on Internet of Things, Big Data, and Security (IoTBDS 2017) **2017**

- Program Committee Member, International Conference on Internet of Things, Big Data, and Security (IoTBDS 2016) **2016**

- Program Committee Member, Big Data and Analytics Education Conference (BDA EdCon European Track) (ECE 2015) **2015**

- Program Committee Member, IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (IEEE AEECT 2013) **2013**

- Program Committee Member, International Conference on Electronics, Computer, and Computation (ICECCO 2013) **2013**

- Program Committee Member, the ACM International Conference on Bioinformatics and Computational Biology (ACM BCB 2012) **2012**

- Program Committee Member, IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (IEEE AEECT 2011) **2011**

- Program Committee Member, the ACM International Conference on Bioinformatics and Computational Biology (ACM BCB 2010) **2010**

- Program Committee Member, the 24th International Symposium on Computer and Information Sciences (ISCIS 2009) **2009**

- Sponsored Member, The American Association for the Advancement of Science (AAAS) **2008 - 2010**

- Assistant to the Editor in Chief, ACM TODS Journal **2008**

- Reviewer for Bioinformatics Journal, Journal of Recent Patents on Computer Science, SIGMOD, SIAM, VLDB, SSDBM, ADVIS, ICDT, PODS, and ISCIS **2004 - Present**