COMP90050 Project Candidate Topics and Example Papers

Here we list some topics to help those who do not know where to start. Groups can use these topics as starting points to explore cutting-edge database technologies, and find the materials on which you feel most excited. The papers listed under each topic can help you know more about related areas, and you can use them (especially their references) as keys to open the door of a huge literature for each topic.

For these example papers, feel free to include or exclude them in your group presentation and final report. When preparing your presentation and final report, you may find it very helpful to first obtain a big picture of your chosen topic by searching and reading a high-quality survey on the topic (e.g., those published in ACM Computing Surveys).

*Please note that crossed-out topics are already assigned to groups.

Possible Topics

1. ~~Shared Nothing Clusters and Databases~~
2. ~~Solid State Drives and Databases~~
3. ~~Main-Memory Data Management~~
4. ~~Transaction Processing~~
5. ~~Concurrency in Data Management~~
6. ~~Data Storage~~
7. ~~Privacy in Data Management~~
8. ~~Data Integration~~
9. ~~Data Cleaning and Quality Management~~
10. ~~Security in Data Management~~
11. ~~Graph Data Management~~
12. ~~Sensor Network Data Management~~
13. ~~Semantic Data Management~~
14. ~~Spatial Data Management~~
15. ~~Social Data Management~~
16. ~~Location based Service Data Management~~
17. ~~Temporal Data Management~~
18. ~~Trajectory Data Management~~
19. ~~Scientific Data Management~~
20. ~~Multimedia Data Management~~
21. ~~Textual Data Management~~
22. ~~RDF Data Management~~
23. ~~XML Data Management~~
24. ~~Top k Query Processing~~
25. ~~Similarity Query Processing~~
26. ~~Recommendation Systems~~
27. ~~Crowdsourcing and Databases~~
28. ~~NoSQL Databases~~
29. ~~Distributed Indexes and Query Processing~~
30. ~~Uncertain Data Processing~~
31. ~~Streaming Data Processing~~
32. ~~Interactive Data Exploration and Analytics~~
33. ~~Large-Scale Data Analytics and Distributed Machine Learning~~
34. ~~Urban Computing~~
35. ~~User Behavior Modeling~~

Example Papers

Database Systems on Modern Hardware and Architecture

1. **Shared Nothing Clusters and Databases**
   a. Sakr, Sherif, Anna Liu, and Ayman G. Fayoumi. "The family of MapReduce and large scale data processing systems." ACM Computing Surveys (CSUR) 46.1 (2013):11.
   b. Li, F. E. N. G., et al. "Distributed data management using mapreduce." ACM Computing Survey (2013).
   c. Low, Yucheng, et al. "Graphlab: A new framework for parallel machine learning." arXiv preprint arXiv:1006.4990 (2010).
   d. Low, Yucheng, et al. "Distributed GraphLab: a framework for machine learning and data mining in the cloud." Proceedings of the VLDB Endowment 5.8 (2012): 716-727.
   e. Mone, Gregory. "Beyond Hadoop." Communications of the ACM 56.1 (2013): 22-24.

2. **Solid State Drives and Databases**
   a. Do, Jaeyoung, et al. "Turbocharging DBMS buffer pool using SSDs." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.
   b. Lee, Sang-Won, Bongki Moon, and Chanik Park. "Advances in flash memory SSD technology for enterprise database applications." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.
   c. Roh, Hongchan, et al. "B+-tree index optimization by exploiting internal parallelism of flash-based solid state drives." Proceedings of the VLDB Endowment 5.4 (2011): 286-297

d. Canim , Mustafa, et al. "SSD bufferpool extensions for database systems."
    Proceedings of the VLDB Endowment 3.1-2 (2010): 1435-1446.

e. Sadoghi, Mohammad, et al. "Making updates disk-I/O friendly using SSDs." Proceedings of the VLDB Endowment 6.11 (2013): 997-1008.

3. Main-Memory Data Management

a. Li, Yinan, and Jignesh M. Patel. "BitWeaving: Fast scans for main memory data processing." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

b. Nobari, Sadegh, et al. "TOUCH: in-memory spatial join by hierarchical data-oriented partitioning." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

c. Ren, Kun, Alexander Thomson, and Daniel J. Abadi. "Lightweight locking for main memory database systems." Proceedings of the 39th international conference on Very Large Data Bases. VLDB Endowment, 2012.

d. Heimel, Max, et al. "Hardware-oblivious parallelism for in-memory column-stores." Proceedings of the VLDB Endowment 6.9 (2013): 709-720.

e. Albutiu, Martina-Cezara, Alfons Kemper, and Thomas Neumann. "Massively parallel sort-merge joins in main memory multi-core database systems." Proceedings of the VLDB Endowment 5.10 (2012): 1064-1075.

Advanced Issues in Database Management

4. Transaction Processing

a. Pang, Gene, et al. "PLANET: making progress with commit processing in unpredictable environments." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.

b. Faleiro, Jose M., Alexander Thomson, and Daniel J. Abadi. "Lazy evaluation of transactions in database systems." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.

c. Jung, Hyungsoo, et al. "A scalable lock manager for multicores." ACM Transactions on Database Systems (TODS) 39.4 (2014): 29.

d. Harizopoulos, Stavros, et al. "OLTP through the looking glass, and what we found there." Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008.

e. Larus, James, and Christos Kozyrakis. "Transactional memory."Communications of the ACM 51.7 (2008): 80-88.

5. Concurrency in Data Management

a. Mozafari, Barzan, et al. "Performance and resource modeling in highly-concurrent oltp workloads." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

b. Duggan, Jennie, et al. "Performance prediction for concurrent database workloads." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

c. Jones, Evan PC, Daniel J. Abadi, and Samuel Madden. "Low overhead concurrency control for partitioned main memory databases." Proceedings of the2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.

d. Psaroudakis, Iraklis, Manos Athanassoulis, and Anastasia Ailamaki. "Sharing data and work across concurrent analytical queries." Proceedings of the VLDB Endowment 6.9 (2013): 637-648.

 e. Wu, Wentao, et al. "Towards predicting query execution time for concurrent and dynamic database workloads." Proceedings of the VLDB Endowment 6.10 (2013): 925-936.

6. **Data Storage**

 a. Le, Wangchao, et al. "Optimal splitters for temporal and multi-version databases." Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. ACM, 2013.

 b. Bornea, Mihaela A., et al. "Building an efficient RDF store over a relational database." Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. ACM, 2013.

 c. Alagiannis, Ioannis, Stratos Idreos, and Anastasia Ailamaki. "H2O: a hands-free adaptive store." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.

 d. Sears, Russell, and Raghu Ramakrishnan. "bLSM: a general purpose log structured merge tree." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

 e. Ozmen, Oguzhan, et al. "Workload-aware storage layout for database systems." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.

7. **Privacy in Data Management**

 a. Chen, Shixi, and Shuigeng Zhou. "Recursive mechanism: towards node differential privacy and unrestricted joins." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

 b. Zhang, Jun, et al. "PrivGene: differentially private model fitting using genetic algorithms." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

 c. Mohan, Prashanth, et al. "GUPT: privacy preserving data analysis made easy." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

 d. Kifer, Daniel, and Ashwin Machanavajjhala. "No free lunch in data privacy." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

 e. Ding, Bolin, et al. "Differentially private data cubes: optimizing noise sources and consistency." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

8. **Data Integration**

 a. Halevy, Alon, Anand Rajaraman, and Joann Ordille. "Data integration: the teenage years." Proceedings of the 32nd international conference on Very large data bases. VLDB Endowment, 2006.

 b. Sarma, Anish Das, Xin Luna Dong, and Alon Halevy. "Data integration with dependent sources." Proceedings of the 14th International Conference on Extending Database Technology. ACM, 2011.

 c. Dong, Xin Luna, Barna Saha, and Divesh Srivastava. "Less is more: Selecting sources wisely for integration." Proceedings of the 39th international conference on Very Large Data Bases. VLDB Endowment, 2012.

 d. Calvanese, Diego, et al. "Query processing under glav mappings for relational and graph databases." Proceedings of the 39th international conference on Very Large Data Bases. VLDB Endowment, 2012.

e. Li, Xian, et al. "Truth finding on the deep web: is the problem solved?." Proceedings of the 39th international conference on Very Large Data Bases. VLDB Endowment, 2012.

9. **Data Cleaning and Quality Management**
   a. Dallachiesa, Michele, et al. "NADEEF: a commodity data cleaning system." Proceedings of the 2013 international conference on Management of data. ACM, 2013.
   b. Yakout, Mohamed, Laure Berti-Équille, and Ahmed K. Elmagarmid. "Don't be SCAREd: use SCalable Automatic REpairing with maximal likelihood and bounded changes." Proceedings of the 2013 international conference on Management of data. ACM, 2013.
   c. Cao, Yang, Wenfei Fan, and Wenyuan Yu. "Determining the relative accuracy of attributes." Proceedings of the 2013 international conference on Management of data. ACM, 2013.
   d. Meliou, Alexandra, et al. "Tracing data errors with view-conditioned causality." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.
   e. Chen, Haiquan, et al. "Leveraging spatio-temporal redundancy for RFID data cleansing." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.

10. **Security in Data Management**
    a. Marczak, William R., et al. "Secureblox: customizable secure distributed data processing." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.
    b. Kozak, Stepan, and Pavel Zezula. "Efficiency and security in similarity cloud services." Proceedings of the VLDB Endowment 6.12 (2013): 1450-1455.
    c. Tu, Stephen, et al. "Processing analytical queries over encrypted data." Proceedings of the 39th international conference on Very Large Data Bases. VLDB Endowment, 2013.
    d. Canim, Mustafa, et al. "Building disclosure risk aware query optimizers for relational databases." Proceedings of the VLDB Endowment 3.1-2 (2010): 13-24.
    e. Allard, Tristan, et al. "Secure personal data servers: a vision paper." Proceedings of the VLDB Endowment 3.1-2 (2010): 25-35.

Semi-Structural and Spatiotemporal Data Management

11. **Graph Data Management**
    a. Angles, Renzo, and Claudio Gutierrez. "Survey of graph database models." ACM Computing Surveys (CSUR) 40.1 (2008): 1.
    b. Khan, Arijit, et al. "Neighborhood based fast graph search in large networks." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.
    c. Mondal, Jayanta, and Amol Deshpande. "Managing large dynamic graphs efficiently." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.
    d. Yang, Shengqi, et al. "Towards effective partition management for large graphs." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

e. Akiba, Takuya, Yoichi Iwata, and Yuichi Yoshida. "Fast exact shortest-path distance queries on large networks by pruned landmark labeling." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

12. Sensor Network Data Management

a. Huang, Zengfeng, et al. "Sampling based algorithms for quantile computation in sensor networks." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

b. Stern, Mirco, Klemens Böhm, and Erik Buchmann. "Processing continuous join queries in sensor networks: a filtering approach." Proceedings of the2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.

c. Giatrakos, Nikos, et al. "TACO: tunable approximate computation of outliers in wireless sensor networks." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.

d. Gandhi, Sorabh, et al. "Gamps: Compressing multi sensor data by grouping and amplitude scaling." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

e. Franke, Conny, and Michael Gertz. "Orden: outlier region detection and exploration in sensor networks." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

13. Semantic Data Management

a. Yakout, Mohamed, et al. "Infogather: entity augmentation and attribute discovery by holistic matching with web tables." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

b. Li, Guoliang, Dong Deng, and Jianhua Feng. "Faerie: efficient filtering algorithms for approximate dictionary-based entity extraction." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

c. Chen, Zhaoqi, Dmitri V. Kalashnikov, and Sharad Mehrotra. "Exploiting context analysis for combining multiple entity resolution systems." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

d. Whang, Steven Euijong, et al. "Entity resolution with iterative blocking." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

e. Arasu, Arvind, and Raghav Kaushik. "A grammar-based entity representation framework for data cleaning." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

14. Spatial Data Managemnt

a. Long, Cheng, et al. "Collective spatial keyword queries: a distance owner-driven approach." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

b. Nobari, Sadegh, et al. "TOUCH: in-memory spatial join by hierarchical data-oriented partitioning." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

c. Samet, Hanan, Jagan Sankaranarayanan, and Michael Auerbach. "Indexing methods for moving object databases: games and other applications." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

d. Lu, Jiaheng, Ying Lu, and Gao Cong. "Reverse spatial and textual k nearest neighbor search." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

    e. Cao, Xin, et al. "Collective spatial keyword querying." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

15. Social Data Management

    a. Mishne, Gilad, et al. "Fast data in the era of big data: Twitter's real-time related query suggestion architecture." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

    b. Borthakur, Dhruba, et al. "Apache Hadoop goes realtime at Facebook." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

    c. Abraham, Lior, et al. "Scuba: diving into data at facebook." Proceedings of the VLDB Endowment 6.11 (2013): 1057-1067.

    d. Curtiss, Michael, et al. "Unicorn: a system for searching the social graph." Proceedings of the VLDB Endowment 6.11 (2013): 1150-1161.

    e. Papagelis, Manos, Gautam Das, and Nick Koudas. "Sampling online social networks." Knowledge and Data Engineering, IEEE Transactions on 25.3 (2013): 662-676.

16. Location based Service Data Management

    a. Götz, Michaela, Suman Nath, and Johannes Gehrke. "Maskit: privately releasing user context streams for personalized mobile applications." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

    b. Hu, Haibo, et al. "Authenticating location-based services without compromising location privacy." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

    c. Thomsen, Jeppe Rishede, Man Lung Yiu, and Christian S. Jensen. "Effective caching of shortest paths for location-based services." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

    d. Stach, Christoph, and Bernhard Mitschang. "Privacy Management for Mobile Platforms--A Review of Concepts and Approaches." Mobile Data Management (MDM), 2013 IEEE 14th International Conference on. Vol. 1. IEEE, 2013.

    e. Yang, Jin, et al. "Energy-Efficient Collaborative Query Processing Framework for Mobile Sensing Services." Mobile Data Management (MDM), 2013 IEEE 14th International Conference on. Vol. 1. IEEE, 2013.

17. Temporal Data Management

    a. Kaufmann, Martin, et al. "Timeline index: A unified data structure for processing queries on temporal data in SAP HANA." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

    b. Le, Wangchao, et al. "Optimal splitters for temporal and multi-version databases." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

    c. Wang, Peng, Haixun Wang, and Wei Wang. "Finding semantics in time series." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

    d. Mueen, Abdullah, Suman Nath, and Jie Liu. "Fast approximate correlation for massive time-series data." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.

    e. Moon, Hyun Jin, Carlo A. Curino, and Carlo Zaniolo. "Scalable architecture and query optimization for transaction-time DBs with evolving schemas." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.

18. **Trajectory Data Management**
    a. Long, Cheng, Raymond Chi-Wing Wong, and H. V. Jagadish. "Direction-preserving trajectory simplification." Proceedings of the VLDB Endowment 6.10 (2013): 949-960.
    b. Luo, Wuman, et al. "Finding time period-based most frequent path in big trajectory data." Proceedings of the 2013 international conference on Management of data. ACM, 2013.
    c. Su, Han, et al. "Calibrating trajectory data for similarity-based analysis." Proceedings of the 2013 international conference on Management of data. ACM, 2013.
    d. Chen, Zaiben, et al. "Searching trajectories by locations: an efficiency study." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.
    e. Zheng, Yu, et al. "Urban Computing: Concepts, Methodologies, and Applications." ACM transactions on Intelligent Systems and Technology (2014).

19. **Scientific Data Management**
    a. Stougiannis, Alexandros, et al. "Data-driven neuroscience: enabling breakthroughs via innovative data management." Proceedings of the 2013 international conference on Management of data. ACM, 2013.
    b. Yang, Xiaochun, Honglei Liu, and Bin Wang. "ALAE: accelerating local alignment with affine gap exactly in biosequence databases." Proceedings of the VLDB Endowment 5.11 (2012): 1507-1518.
    c. Kementsietsidis, Anastasios, et al. "Scalable multi-query optimization for exploratory queries over federated scientific databases." Proceedings of the VLDB Endowment 1.1 (2008): 16-27.
    d. Soroush, Emad, and Magdalena Balazinska. "Time travel in a scientific array database." Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE, 2013.
    e. Wu, Eugene, Samuel Madden, and Michael Stonebraker. "Subzero: a fine-grained lineage system for scientific databases." Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE, 2013.

20. **Multimedia Data Management**
    a. Sun, Xinghai, et al. "Indexing billions of images for sketch-based retrieval." Proceedings of the 21st ACM international conference on Multimedia. ACM, 2013.
    b. Liu, Wu, et al. "LAVES: an instant mobile video search system based on layered audiovideo indexing." Proceedings of the 21st ACM international conference on Multimedia. ACM, 2013.
    c. Liu, Wu, et al. "Listen, look, and gotcha: instant video search with mobile phones by layered audio-video indexing." Proceedings of the 21st ACM international conference on Multimedia. ACM, 2013.
    d. Cao, Liangliang, et al. "Submodular video hashing: a unified framework towards video pooling and indexing." Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012.
    e. Ma, Zhigang, et al. "Exploiting the entire feature space with sparsity for automatic image annotation." Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011.

21. **Textual Data Management**
    a. Yang, Xiaochun, et al. "Improving regular-expression matching on strings using negative factors." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

b. Lu, Jiaheng, et al. "String similarity measures and joins with synonyms." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

c. Wu, Wentao, et al. "Probase: A probabilistic taxonomy for text understanding." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

d. Ding, Bolin, et al. "Optimizing index for taxonomy keyword search." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

e. Yang, Xiaochun, Bin Wang, and Chen Li. "Cost-based variable-length-gram selection for string collections to support approximate queries efficiently." Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008.

22. RDF Data Management

a. Gurajada, Sairam, et al. "TriAD: A distributed shared-nothing RDF engine based on asynchronous message passing." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.

b. Bornea, Mihaela A., et al. "Building an efficient RDF store over a relational database." Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. ACM, 2013.

c. Duan, Songyun, et al. "Apples and oranges: a comparison of RDF benchmarks and real RDF datasets." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

d. Lian, Xiang, and Lei Chen. "Efficient query answering in probabilistic rdf graphs."Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

e. Huang, Jiewen, Daniel J. Abadi, and Kun Ren. "Scalable SPARQL querying of large RDF graphs." Proceedings of the VLDB Endowment 4.11 (2011): 1123-1134.

23. XML Data Management

a. Truong, Ba Quan, et al. "MESSIAH: missing element-conscious SLCA nodes search in XML data." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

b. Arenas, Marcelo, et al. "Discovering XSD keys from XML data." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

c. Tao, Yufei, et al. "Nearest keyword search in XML documents." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

d. Taranov, Ilya, et al. "Sedna: native xml database management system (internals overview)." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.

e. Bex, Geert Jan, et al. "Simplifying XML schema: effortless handling of nondeterministic regular expressions." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

Advanced Query Processing and Applications

24. Top-k Queries

a. Yu, Albert, Pankaj K. Agarwal, and Jun Yang. "Processing a large number of continuous preference top-k queries." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

b. Lu, Jiaheng, et al. "Optimal top-k generation of attribute combinations based on ranked lists." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

c. Fraternali, Piero, Davide Martinenghi, and Marco Tagliasacchi. "Top-k bounded diversification." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

d. Mouratidis, Kyriakos, and HweeHwa Pang. "Computing immutable regions for subspace top-k queries." Proceedings of the 39th international conference on Very Large Data Bases. VLDB Endowment, 2012.

e. Ranu, Sayan, and Ambuj K. Singh. "Answering top-k queries over a mixture of attractive and repulsive dimensions." Proceedings of the VLDB Endowment 5.3 (2011): 169-180.

25. Similarity Queries

a. Silva, Yasin N., et al. "Similarity queries: their conceptual evaluation, transformations, and processing." The VLDB Journal—The International Journal on Very Large Data Bases 22.3 (2013): 395-420.

b. Cohen, Sara. "Indexing for subtree similarity-search using edit distance." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

c. Li, Guoliang, Dong Deng, and Jianhua Feng. "A partition-based method for string similarity joins with edit-distance constraints." ACM Transactions on Database Systems (TODS) 38.2 (2013): 9.

d. Wang, Ye, Ahmed Metwally, and Srinivasan Parthasarathy. "Scalable all-pairs similarity search in metric spaces." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.

26. Recommendation Systems

a. Gupta, Pankaj, et al. "Wtf: The who to follow service at twitter." Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013.

b. Davidson, James, et al. "The YouTube video recommendation system." Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010.

c. Ye, Mao, Peifeng Yin, and Wang-Chien Lee. "Location recommendation for location-based social networks." Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2010.

d. Gionis, Aristides, et al. "Customized tour recommendations in urban areas." Proceedings of the 7th ACM international conference on Web search and data mining. ACM, 2014.

e. Ho, Yu-Chieh, Yi-Ting Chiang, and Jane Yung-Jen Hsu. "Who likes it more?: mining worth-recommending items from long tails by modeling relative preference." Proceedings of the 7th ACM international conference on Web search and data mining. ACM, 2014.

Advanced Techniques for Emerging Issues in Databases

27. Crowdsourcing and Databases

a. Doan, Anhai, Raghu Ramakrishnan, and Alon Y. Halevy. "Crowdsourcing systems on the world-wide web." Communications of the ACM 54.4 (2011): 86-96.

b. Marcus, Adam, et al. "Counting with the crowd." Proceedings of the 39th international conference on Very Large Data Bases. VLDB Endowment, 2012.

    c.   Whang, Steven Euijong, Peter Lofgren, and Hector Garcia-Molina. "Question selection for crowd entity resolution." Proceedings of the VLDB Endowment 6.6 (2013): 349-360.

    d.   Kaplan, Haim, et al. "Answering planning queries with the crowd." Proceedings of the VLDB Endowment 6.9 (2013): 697-708.

    e.   Zhang, Chen Jason, et al. "Reducing uncertainty of schema matching via crowdsourcing." Proceedings of the VLDB Endowment 6.9 (2013): 757-768.

28. NoSQL Databases

    a.   Leavitt, Neal. "Will NoSQL databases live up to their promise?." Computer 43.2 (2010):12-14.

    b.   Cattell, Rick. "Scalable SQL and NoSQL data stores." ACM SIGMOD Record 39.4 (2011):12-27.

    c.   Stonebraker, Michael. "SQL databases v. NoSQL databases." Communications of the ACM 53.4 (2010): 10-11.

    d.   Qiao, Lin, et al. "On brewing fresh espresso: LinkedIn's distributed data serving platform." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

    e.   Sivasubramanian, Swaminathan. "Amazon dynamoDB: a seamlessly scalable nonrelational database service." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

29. Distributed Indexes and Query Processing

    a.   Song, Jingkuan, et al. "Inter-media hashing for large-scale retrieval from heterogeneous data sources." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

    b.   Ghoting, Amol, and Konstantin Makarychev. "Serial and parallel methods for i/o efficient suffix tree construction." Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009.

    c.   Shute, Jeff, et al. "F1: A distributed SQL database that scales." Proceedings of the VLDB Endowment 6.11 (2013): 1068-1079.

    d.   Manshadi, Faraz Makari, et al. "A distributed algorithm for large-scale generalized matching." Proceedings of the VLDB Endowment 6.9 (2013): 613-624.

    e.   Aguilera, Marcos K., Wojciech Golab, and Mehul A. Shah. "A practical scalable distributed b-tree." Proceedings of the VLDB Endowment 1.1 (2008): 598-609.

30. Uncertain Data Processing

    a.   Tran, Thanh TL, et al. "Supporting user-defined functions on uncertain data." Proceedings of the VLDB Endowment 6.6 (2013): 469-480.

    b.   Lian, Xiang, and Lei Chen. "A generic framework for handling uncertain data with local correlations." Proceedings of the VLDB Endowment 4.1 (2010): 12-21.

    c.   Xu, Jie, Dmitri V. Kalashnikov, and Sharad Mehrotra. "Efficient summarization framework for multi-attribute uncertain data." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.

    d.   Parchas, Panos, et al. "The pursuit of a good possible world: extracting representative instances of uncertain graphs." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.

    e.   Rekatsinas, Theodoros, Amol Deshpande, and Lise Getoor. "Local structure and determinism in probabilistic databases." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

31. Streaming Data Processing

a. Ananthanarayanan, Rajagopal, et al. "Photon: Fault-tolerant and scalable joining of continuous data streams." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

b. Wang, Di, et al. "Utility-maximizing event stream suppression." Proceedings of the 2013 international conference on Management of data. ACM, 2013.

c. Mozafari, Barzan, Kai Zeng, and Carlo Zaniolo. "High-performance complex event processing over xml streams." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

d. Giatrakos, Nikos, et al. "Prediction-based geometric monitoring over distributed data streams." Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012.

e. Teubner, Jens, and Rene Mueller. "How soccer players would do stream joins." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.

Mining Knowledge from Database

32. Interactive Data Exploration and Analytics

a. Idreos, Stratos, Olga Papaemmanouil, and Surajit Chaudhuri. "Overview of data exploration techniques." Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. ACM, 2015.

b. Emin Aksehirli, Bart Goethals, and Emmanuel Müller, "Visual Interactive Neighborhood Mining on High Dimensional Data." IDEA Workshop on KDD 2015

c. Dou, Wenwen, et al. "Leadline: Interactive visual analysis of text data through event identification and exploration." Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on. IEEE, 2012.

d. Subhajit Das, Andrea McCarter, Joe Minieri, Nandita Damaraju, Sriram Padmanabhan and Duen Horng Chau "ISPARK: Interactive Visual Analytics for Fire Incidents and Station Placement." IDEA Workshop on KDD 2015

e. Madden, Samuel. "Interactive data analytics: the new frontier." Proceedings of the Sixth ACM Symposium on Cloud Computing. ACM, 2015.

33. Large-Scale Data Analytics and Distributed Machine Learning

a. Shanahan, James G., and Laing Dai. "Large scale distributed data science using apache spark." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

b. Low, Yucheng, et al. "Distributed GraphLab: a framework for machine learning and data mining in the cloud." Proceedings of the VLDB Endowment 5.8 (2012): 716-727.

c. Kraska, Tim, et al. "MLbase: A Distributed Machine-learning System." CIDR. Vol. 1. 2013.

d. Sparks, Evan R., et al. "MLI: An API for distributed machine learning." Data Mining (ICDM), 2013 IEEE 13th International Conference on. IEEE, 2013.

e. Li, Mu, et al. "Communication efficient distributed machine learning with the parameter server." Advances in Neural Information Processing Systems. 2014.

34. Urban Computing

a. Zheng, Yu, et al. "Urban computing: concepts, methodologies, and applications." ACM Transactions on Intelligent Systems and Technology (TIST) 5.3 (2014): 38.

b. Zheng, Yu, et al. "Urban computing with taxicabs." Proceedings of the 13th international conference on Ubiquitous computing. ACM, 2011.

   c. Jiang, Shan, et al. "A review of urban computing for mobile phone traces: current methods, challenges and opportunities." Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing. ACM, 2013.

   d. Minh Tuan Doan, et al. "Profiling Pedestrian Activity Patterns in a Dynamic Urban Environment." UrbComp 2015

   e. Dheeraj Kumar, et al. "A Scalable Framework for Clustering Vehicle Trajectories in a Dense Road Network." UrbComp 2015

35. User Behavior Modeling

   a. Morita, Masahiro, and Yoichi Shinoda. "Information filtering based on user behavior analysis and best match text retrieval." Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. Springer-Verlag New York, Inc., 1994.

   b. Yin, Hongzhi, et al. "A temporal context-aware model for user behavior modeling in social media systems." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.

   c. Teevan, Jaime, Susan T. Dumais, and Daniel J. Liebling. "To personalize or not to personalize: modeling queries with variation in user intent." Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2008.

36. Entity Recognition

   a. Li, Chenliang, et al. "Twiner: named entity recognition in targeted twitter stream." Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2012.

   b. Ling, Xiao, and Daniel S. Weld. "Fine-Grained Entity Recognition." AAAI. 2012.

   c. Ren, Xiang, et al. "Clustype: Effective entity recognition and typing by relation phrase-based clustering." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

   d. Nothman, Joel, et al. "Learning multilingual named entity recognition from Wikipedia." Artificial Intelligence 194 (2013): 151-175.

   e. Zhang, Lei, and Bing Liu. "Aspect and entity extraction for opinion mining." Data mining and knowledge discovery for big data. Springer Berlin Heidelberg, 2014. 1-40.

37. Time Series Management and Mining

   a. Sakurai, Yasushi, Yasuko Matsubara, and Christos Faloutsos. "Mining and Forecasting of Big Time-series Data." Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. ACM, 2015.

   b. Cai, Yongjie, et al. "Facets: Fast Comprehensive Mining of Coevolving High-order Time Series." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

   c. Moskovitch, Robert, and Yuval Shahar. "Classification of multivariate time series via temporal abstraction and time intervals mining." Knowledge and Information Systems 45.1 (2015): 35-74.

   d. Esling, Philippe, and Carlos Agon. "Time-series data mining." ACM Computing Surveys (CSUR) 45.1 (2012): 12.

38. Sentiment Discovery and Analysis

   a. Feldman, Ronen. "Techniques and applications for sentiment analysis." Communications of the ACM 56.4 (2013): 82-89.

b. Liu, Bing, and Lei Zhang. "A survey of opinion mining and sentiment analysis." Mining text data. Springer US, 2012. 415-463.

c. Kontopoulos, Efstratios, et al. "Ontology-based sentiment analysis of twitter posts." Expert systems with applications 40.10 (2013): 4065-4074.

d. Cambria, Erik, et al. "AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis." AAAI. 2015.

39. Anomaly Detection

a. Li, Weixin, Vijay Mahadevan, and Nuno Vasconcelos. "Anomaly detection and localization in crowded scenes." Pattern Analysis and Machine Intelligence, IEEE Transactions on 36.1 (2014): 18-32.

b. Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection for discrete sequences: A survey." Knowledge and Data Engineering, IEEE Transactions on 24.5 (2012): 823-839.

c. Savage, David, et al. "Anomaly detection in online social networks." Social Networks 39 (2014): 62-70.

d. Bhuyan, Monowar H., Dhruba Kumar Bhattacharyya, and Jugal Kumar Kalita. "Network anomaly detection: methods, systems and tools." Communications Surveys & Tutorials, IEEE 16.1 (2014): 303-336.

40. Topic Modeling

a. Hu, Yuening, et al. "Interactive topic modeling." Machine learning 95.3 (2014): 423-469.

b. Blei, David M. "Probabilistic topic models." Communications of the ACM 55.4 (2012): 77-84.

c. Tang, Jian, et al. "Understanding the limiting factors of topic modeling via posterior contraction analysis." Proceedings of The 31st International Conference on Machine Learning. 2014.

d. Gretarsson, Brynjar, et al. "Topicnets: Visual analysis of large text corpora with topic modeling." ACM Transactions on Intelligent Systems and Technology (TIST) 3.2 (2012): 23.

41. Online Social Network Analysis

a. Jiang, Jing, et al. "Understanding latent interactions in online social networks." ACM Transactions on the Web (TWEB) 7.4 (2013): 18.

b. Guille, Adrien, et al. "Information diffusion in online social networks: A survey." ACM SIGMOD Record 42.2 (2013): 17-28.

c. Kramer, Adam DI, Jamie E. Guillory, and Jeffrey T. Hancock. "Experimental evidence of massive-scale emotional contagion through social networks." Proceedings of the National Academy of Sciences 111.24 (2014): 8788-8790.

d. Liu, Xingjie, et al. "Event-based social networks: linking the online and offline social worlds." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.

42. Web Mining

a. Berendt, Bettina. "More than modelling and hiding: towards a comprehensive view of Web mining and privacy." Data Mining and Knowledge Discovery 24.3 (2012): 697-737.

b. Kosala, Raymond, and Hendrik Blockeel. "Web mining research: A survey." ACM Sigkdd Explorations Newsletter 2.1 (2000): 1-15.

c. Eirinaki, Magdalini, and Michalis Vazirgiannis. "Web mining for web personalization." ACM Transactions on Internet Technology (TOIT) 3.1 (2003): 1-27.

d. Mulvenna, Maurice D., Sarabjot S. Anand, and Alex G. Büchner. "Personalization on the Net using Web mining: introduction." Communications of the ACM 43.8 (2000): 122-125.

43. Knowledge Graph and Knowledge Base

a. Hoffart, Johannes, et al. "YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia." Artificial Intelligence 194 (2013): 28-61.

b. Pujara, Jay, et al. "Knowledge graph identification." The Semantic Web–ISWC 2013. Springer Berlin Heidelberg, 2013. 542-557.

c. Bollacker, Kurt, et al. "Freebase: a collaboratively created graph database for structuring human knowledge." Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008.

d. Chen, Yun-Nung, et al. "Matrix factorization with knowledge graph propagation for unsupervised spoken language understanding." Proceedings of ACL-IJCNLP (2015).

e. Yahya, Mohamed, et al. "Relationship queries on extended knowledge graphs." Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. ACM, 2016.

f. Pound, Jeffrey, et al. "Interpreting keyword queries over web knowledge bases." Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012.

44. Security and Privacy for Data Mining

a. Aggarwal, Charu C., and S. Yu Philip. A general survey of privacy-preserving data mining models and algorithms. Springer US, 2008.

b. Clifton, Chris, et al. "Tools for privacy preserving distributed data mining." ACM Sigkdd Explorations Newsletter 4.2 (2002): 28-34.

c. Chen, Feng, et al. "Data mining for the internet of things: literature review and challenges." International Journal of Distributed Sensor Networks 2015 (2015): 12.

d. Payne, Dinah, Brett JL Landry, and Matthew D. Dean. "Data Mining and Privacy: An Initial Attempt at a Comprehensive Code of Conduct for Online Business." Communications of the Association for Information Systems 37.1 (2015): 34.

e. Machanavajjhala, Ashwin, and Daniel Kifer. "Designing statistical privacy for your data." Communications of the ACM 58.3 (2015): 58-67.

45. Dimensionality Reduction and Feature Selection

a. Chizi, Barak, and Oded Maimon. "Dimension reduction and feature selection." Data mining and knowledge discovery handbook. Springer US, 2009. 83-100.

b. Guyon, Isabelle, and André Elisseeff. "An introduction to variable and feature selection." The Journal of Machine Learning Research 3 (2003): 1157-1182.

c. Zhao, Zheng, et al. "On similarity preserving feature selection." Knowledge and Data Engineering, IEEE Transactions on 25.3 (2013): 619-632.

d. Cohen, Michael B., et al. "Dimensionality reduction for k-means clustering and low rank approximation." Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing. ACM, 2015.

e. Xu, Chang, et al. "Large-margin weakly supervised dimensionality reduction." Proceedings of the 31st International Conference on Machine Learning (ICML-14). 2014.

46. Medical Data Management and Mining

    a. Aggarwal, Charu C., ed. Managing and mining sensor data. Springer Science & Business Media, 2013.

    b. Fan, Kai, et al. "Hierarchical Graph-Coupled HMMs for Heterogeneous Personalized Health Data." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

    c. Caballero Barajas, Karla L., and Ram Akella. "Dynamically Modeling Patient's Health State from Electronic Medical Records: A Time Series Approach." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

    d. Wang, Zheng, et al. "Dynamic Poisson Autoregression for Influenza-Like-Illness Case Count Prediction." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

47. Similarity and Hashing

    a. Wang, Jingdong, et al. "Hashing for similarity search: A survey." arXiv preprint arXiv:1408.2927 (2014).

    b. Wang, Jun, et al. "Learning to Hash for Indexing Big Data—A Survey." Proceedings of the IEEE 104.1 (2016): 34-57.

    c. Zhang, Dell, et al. "Self-taught hashing for fast similarity search." Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2010.

    d. Wang, Jingjing, and Chen Lin. "MapReduce based personalized locality sensitive hashing for similarity joins on large scale data." Computational intelligence and neuroscience 2015 (2015): 37.

48. Business, Sales, Marketing, and Advertising Driven Data Analytics

    a. Xu, Jian, et al. "Smart Pacing for Effective Online Ad Campaign Optimization." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

    b. Shen, Jianqiang, Sahin Cem Geyik, and Ali Dasdan. "Effective Audience Extension in Online Advertising." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

    c. Ikonomovska, Elena, Sina Jafarpour, and Ali Dasdan. "Real-Time Bid Prediction using Thompson Sampling-Based Expert Selection." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

    d. Li, Cheng, et al. "Click-through prediction for advertising in twitter timeline." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

49. E-Commerce Data Management

    a. Zhong, Wenliang, et al. "Stock Constrained Recommendation in Tmall." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

    b. Mihajlo Grbovic, Vladan Radosavljevic, Nemanja Djuric, Narayan Bhamidipati, "E-commerce in your Inbox: Cross-domain product recommendations at scale." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

    c.   Wang, Jian, and Yi Zhang. "Opportunity model for e-commerce recommendation: right product; right time." Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2013.

    d.   Jiang, Peng, et al. "Life-stage Prediction for Product Recommendation in E-commerce." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

50. Query Understanding

    a.   Demartini, Gianluca, et al. "CrowdQ: Crowdsourced Query Understanding." CIDR. 2013.

    b.   Goyal, Amit, et al. "Query Understanding for Search on All Devices at WSDM 2016." Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. ACM, 2016.

    c.   Kravi, Elad, et al. "Searcher in a strange land: Understanding web search from familiar and unfamiliar locations." Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015.

    d.   Miliaraki, Iris, Roi Blanco, and Mounia Lalmas. "From Selena Gomez to Marlon Brando: Understanding Explorative Entity Search." Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2015.

    e.   Hasanuzzaman, Mohammed, et al. "Understanding Temporal Query Intent." Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015.

51. Scaling-up Methods for Big Data

    a.   Landset, Sara, et al. "A survey of open source tools for machine learning with big data in the Hadoop ecosystem." Journal of Big Data 2.1 (2015): 1-36.

    b.   Singh, Dilpreet, and Chandan K. Reddy. "A survey on platforms for big data analytics." Journal of big data 2.1 (2015): 1-20.

    c.   Abadi, Daniel, et al. "The beckman report on database research." Communications of the ACM 59.2 (2016): 92-99.

    d.   Lin, Jimmy, and Dmitriy Ryaboy. "Scaling big data mining infrastructure: the twitter experience." ACM SIGKDD Explorations Newsletter 14.2 (2013): 6-19.

    e.   Chen, CL Philip, and Chun-Yang Zhang. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data." Information Sciences 275 (2014): 314-347.

52. Behavior Analysis and Personalization

    a.   Hsieh, Cheng-Kang, et al. "Immersive Recommendation: News and Event Recommendations Using Personal Digital Traces." Proceedings of the 25th ACM International Conference on World Wide Web. ACM. 2016.

    b.   White, Ryen W., et al. "Enhancing personalized search by mining and modeling task behavior." Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013.

    c.   Aditya Pal, Amaç Herdağdelen, Sourav Chatterji, Sumit Taank and Deepayan Chakrabarti, "Discovery of Topical Authorities in Instagram." Proceedings of the 25th ACM International Conference on World Wide Web. ACM. 2016.

    d.   Bennett, Paul N., et al. "Modeling the impact of short-and long-term behavior on search personalization." Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2012.

   e. Milad Shokouhi, Umut Ozertem and Nick Craswell, "Did You Say U2 or Youtube? Inferring Implicit Transcripts from Voice Search Logs." Proceedings of the 25th ACM International Conference on World Wide Web. ACM. 2016.

53. Statistical Techniques for Big Data

   a. Cohen, Jeffrey, et al. "MAD skills: new analysis practices for big data." Proceedings of the VLDB Endowment 2.2 (2009): 1481-1492.

   b. LaValle, Steve, et al. "Big data, analytics and the path from insights to value." MIT sloan management review 52.2 (2011): 21.

   c. Ott, R. L., and Micheal Longnecker. An introduction to statistical methods and data analysis. Nelson Education, 2015.

   d. Lazer, David, et al. "The parable of Google Flu: traps in big data analysis." Science 343.14 March (2014).

   e. Armitage, Peter, Geoffrey Berry, and John Nigel Scott Matthews. Statistical methods in medical research. John Wiley & Sons, 2008.

   f. Miner, Gary, Robert Nisbet, and John Elder IV. Handbook of statistical analysis and data mining applications. Academic Press, 2009.

54. Large-scale Optimization and Learning

   a. Raina, Rajat, Anand Madhavan, and Andrew Y. Ng. "Large-scale deep unsupervised learning using graphics processors." Proceedings of the 26th annual international conference on machine learning. ACM, 2009.

   b. Akata, Zeynep, et al. "Good practice in large-scale learning for image classification." Pattern Analysis and Machine Intelligence, IEEE Trans. on 36.3 (2014): 507-520.

   c. Bousquet, Olivier, and Léon Bottou. "The tradeoffs of large scale learning." Advances in neural information processing systems. 2008.

   d. Rao, R. Venkata, Vimal J. Savsani, and D. P. Vakharia. "Teaching–learning-based optimization: an optimization method for continuous non-linear large scale problems." Information Sciences 183.1 (2012): 1-15.

   e. Dean, Jeffrey, et al. "Large scale distributed deep networks." Advances in Neural Information Processing Systems. 2012.

55. Transportation Data Analysis

   a. Du, Fan, et al. "Metroviz: Visual analysis of public transportation data." arXiv preprint arXiv:1507.05215 (2015).

   b. Xiao, Qin. "Big Data analysis: Intelligent Transportation Development Engine." Urban Transportation & Construction 1.1 (2014): 11-13.

   c. Waldheim, Nicole, Elizabeth Wemple, and Joseph Fish. Applying Safety Data and Analysis to Performance-Based Transportation Planning. No. FHWA-SA-15-089. 2015.

   d. Nakamura, Toshiyuki, et al. "Urban Public Transport Mileage Cards: Analysis of Their Potential with Smart Card Data and an SP Survey." Transportation Research Board 95th Annual Meeting. No. 16-4645. 2016.

   e. Becker, Richard, et al. "Human mobility characterization from cellular network data." Communications of the ACM 56.1 (2013): 74-82.

   f. Alexander, Lauren P., and Marta C. González. "Assessing the Impact of Real-time Ridesharing on Urban Traffic using Mobile Phone Data." UrbComp Workshop 2015.

56. Text Mining and Sequential Pattern Mining

   a. Kimura, Mitsuki, Atsuhiro Takasu, and Jun Adachi. "FPI: a novel indexing method using frequent patterns for approximate string searches." Proceedings of the Joint EDBT/ICDT 2013 Workshops. ACM, 2013.

b. Mooney, Carl H., and John F. Roddick. "Sequential pattern mining--approaches and algorithms." ACM Computing Surveys (CSUR) 45.2 (2013): 19.

c. Stilo, Giovanni, and Paola Velardi. "Efficient temporal mining of micro-blog texts and its application to event discovery." Data Mining and Knowledge Discovery (2015): 1-31.

d. Gomaa, Wael H., and Aly A. Fahmy. "A survey of text similarity approaches." International Journal of Computer Applications 68.13 (2013).

57. Collaborative Filtering

a. Sarwar, Badrul, et al. "Item-based collaborative filtering recommendation algorithms." Proceedings of the 10th international conference on World Wide Web. ACM, 2001.

b. Shi, Yue, et al. "CLiMF: learning to maximize reciprocal rank with collaborative less-is-more filtering." Proceedings of the sixth ACM conference on Recommender systems. ACM, 2012.

c. Yang, Xiwang, et al. "A survey of collaborative filtering based social recommender systems." Computer Communications 41 (2014): 1-10.

d. Tang, Mingdong, et al. "Location-aware collaborative filtering for QoS-based service recommendation." Web Services (ICWS), 2012 IEEE 19th International Conference on. IEEE, 2012.

e. Paquet, Ulrich, and Noam Koenigstein. "One-class collaborative filtering with random graphs." Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013.

Pervasive and Ubiquitous Computing

58. Sensing the Body

a. Hardegger, Michael, et al. "Enhancing action recognition through simultaneous semantic mapping from body-worn motion sensors." Proceedings of the 2014 ACM International Symposium on Wearable Computers. ACM, 2014.

b. Kawamoto, Ken, Takeshi Tanaka, and Hiroyuki Kuriyama. "Your activity tracker knows when you quit smoking." Proceedings of the 2014 ACM International Symposium on Wearable Computers. ACM, 2014.

c. Aminian, Kamiar, et al. "Gait analysis using shoe-worn inertial sensors: how is foot clearance related to walking speed?." Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014.

d. Abdullah, Saeed, et al. "Towards circadian computing: early to bed and early to rise makes some of us unhealthy and sleep deprived." Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014.

e. Gu, Weixi, et al. "Intelligent sleep stage mining service with smartphones." Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014.

59. Indoor Location and Home Intelligence

a. Chang, Kerry Shih-Ping, Catalina M. Danis, and Robert G. Farrell. "Lunch Line: using public displays and mobile devices to encourage healthy eating in an organization." Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014.

b. Koehler, Christian, et al. "Indoor-alps: an adaptive indoor location prediction system." Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014.

c. Lee, Seungwoo, et al. "Non-obstructive room-level locating system in home environments using activity fingerprints from smartwatch." Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2015.

d. Elhamshary, Moustafa, and Moustafa Youssef. "CheckInside: A fine-grained indoor Location-based social Network." Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014.

e. Viani, Federico, et al. "Wireless architectures for heterogeneous sensing in smart home applications: Concepts and real implementation." Proceedings of the IEEE 101.11 (2013): 2381-2396.

f. Cook, Diane J. "How smart is your home?." Science 335.6076 (2012): 1579-1581.