

Continuous Top-k Query for Graph Streams

Shirui Pan and Xingquan Zhu
Centre for Quantum Computation and Intelligent Systems
Faculty of Engineering and Information Technology
University of Technology, Sydney, NWS 2007, Australia
Shirui.Pan@student.uts.edu.au, Xingquan.Zhu@uts.edu.au

ABSTRACT

In this paper, we propose to query correlated graphs in a data stream scenario, where an algorithm is required to retrieve the top k graphs which are mostly correlated to a query graph q . Due to the dynamic changing nature of the stream data and the inherent complexity of the graph query process, treating graph streams as static datasets is computationally infeasible or ineffective. In the paper, we propose a novel algorithm, Hoe-PGPL, to identify top- k correlated graphs from data stream, by using a sliding window which covers a number of consecutive batches of stream data records. Our theme is to employ Hoeffding bound to discover some potential candidates and use two level candidate checking (one corresponding to the whole sliding window level and one corresponding to the local data batch level) to accurately estimate the correlation of the emerging candidate patterns, without rechecking the historical stream data. Experimental results demonstrate that the proposed algorithm not only achieves good performance in terms of query precision and recall, but also is several times, or even an order of magnitude, more efficient than the straightforward algorithm with respect to the time and the memory consumption. Our method represents the first research endeavor for data stream based top- k correlated graph query.

Categories and Subject Descriptors

H.2.8 [Data Management]: Database Applications—*Data Mining*

General Terms

Algorithm

Keywords

Correlated Graph Query, Graph Stream, Pearson's Correlation Coefficient

1. INTRODUCTION

Correlation mining is a crucial data mining task due to its uniqueness and advantage for revealing underlying dependency between objects. Recent years have witnessed extensive studies on correlation mining in various applications including market transaction databases [6] and quantitative databases [4]. This topic has been recently extended to graph databases, where data records or instances are linked through structural relationships [3, 5].

Correlation between two graphs measures the similarity of the pair's occurrence distributions. Given a graph database and a query graph q , correlated graph search (CGSearch) [3] intends to discover subgraph structures whose Pearson's correlation coefficients with q are larger than a given threshold θ . While discovering this type of correlated subgraphs patterns is crucial in many graph representation scenarios, it requires users to specify a threshold. In practice, it is difficult to decide a proper threshold because a low threshold value may result in a large number of correlated graphs and a high value may result in very few answers. Recently a top- k version of algorithm [5] is proposed to identify top- k correlated subgraphs from a database that have the highest correlation with q .

The above two works [3] [5] are limited to static graph databases. In practice, applications may involve data which constantly change or evolve over time. For instance, E-commerce web sites, such as Amazon, may have thousands of users every day. Each user's browsing history can be represented as a graph and the browsing patterns may evolve from time to time and vary from users to users. Correlated graph analysis from user browsing streams provides web site owners power to understand the users in a dynamic and instantaneous way. Unfortunately, to the best of our knowledge, there is no existing research to query correlated graphs in data streams.

In this paper, we investigate top- k correlated graph query for data streams, with the following identified challenges:

- Each subgraph is a candidate of correlated graph to the query q and there are an exponential number of subgraphs in the graph databases. The challenge is further aggravated as computing the correlation for a candidate involves subgraph isomorphism which is NP-complete.
- The correlation between graphs is constantly changing over stream and recomputing all the correlation is time-consuming.
- The streaming scenario requires the algorithm to return answers in a timely fashion.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.
Copyright 2012 ACM 978-1-4503-1156-4/12/10 ...\$15.00.

A straightforward approach to solve our problem is exhaustive search, which uses a sliding window to scan the stream and computes the correlation by employing CGSearch [3] or TopCor [5] algorithms to query the top- k correlated graphs in each sliding window. While this exhaustive approach can ensure complete and correct results, it is computationally inefficient because the reoccurring query process in each window requires a significant amount of computational costs. In this paper, we propose a Hoeffding bound [2] based algorithm. More specifically, we utilize a Hoeffding bound to discover a set of top- k potential candidate graphs from each data batch. Then we design a *global-local check* scheme with two levels of lists to store candidate information, *i.e.*, potential global list (PG) and potential local lists (PLs). By maintaining and integrating the local lists and the global list carefully, we are able to estimate the true correlation of each candidate with small bias. Experimental results on chemical graph data stream demonstrate that our algorithm is several times more efficient than an exhaustive search method *w.r.t.* the system runtime and the memory consumption, and also achieves good performance in terms of query precision and recall.

2. PRELIMINARIES AND PROBLEM DEFINITION

2.1 Preliminaries

We consider undirected, labeled connected graphs in this paper. A graph is denoted as $g = (V, E, \ell)$, where V is the vertices set, E is the edge set, and ℓ is a labeling function to assign labels to a node or an edge. Given two graphs $g_1 = (V_1, E_1, \ell_1)$ and $g_2 = (V_2, E_2, \ell_2)$, a subgraph isomorphism from g_1 to g_2 is an injective function $f: V_1 \rightarrow V_2$, such that $\forall (u, v) \in E_1$, we have $(f(u), f(v)) \in E_2$, $\ell_1(u) = \ell_2(f(u))$, $\ell_1(v) = \ell_2(f(v))$, $\ell_1(u, v) = \ell_2(f(u), f(v))$. It is known that subgraph isomorphism is an NP-complete problem [1]. A graph g_1 is a subgraph of another graph g_2 ($g_1 \subseteq g_2$) if there is a subgraph isomorphism from g_1 to g_2 .

Given a batch of graphs G_i , the projected dataset with respect to a graph g is denoted by $S_g^{G_i} = \bigcup \{g' | g \subseteq g', g' \in G_i\}$, whose frequency N_g is denoted by $N_g = |S_g^{G_i}|$, where $|\bullet|$ denotes the cardinality of the set \bullet . The joint frequency between two graphs g and q in G_i is the number of graphs in G_i which contains both graph g and q , denoted by $N_{gq} = |S_g^{G_i} \cap S_q^{G_i}|$.

2.2 Problem Definition

Given a query graph q and a continuous graph stream G , we emphasize on discovering top- k correlated graphs between q and G . Because G represents a dynamic graph stream which is constantly changing, we assume that graph data arrive batch by batch, and use a sliding window $D = \{G_{i-w+1}, G_{i-w+2}, \dots, G_i\}$ to denote a consecutive region of the graph stream, where $G_j, i-w+1 \leq j \leq i$ represents a batch of graphs and G_i is the most recent batch of the graph stream. Then our problem is to continuously monitor and report the top- k correlated graphs with the highest Pearson's correlation with q in a sliding window (the most recent w batches). A typical top- k correlated graph query in a data stream with window size $w=3$ is shown in Fig. 1.

Definition 1. Pearson's Correlation Coefficient: Pearson's Correlation Coefficient [6] between two graphs g and q

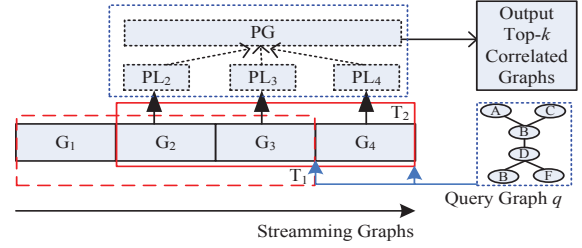


Figure 1: A sliding window based top- k correlated graph query for data streams. At time point T_1 , the sliding window (dashed red rectangle) covers batches G_1 , G_2 , and G_3 . A new batch G_4 arrives at time point T_2 (G_4 is the most recent batch), the sliding window updates to cover G_2 , G_3 , and G_4 (solid red rectangle). We use two levels of candidate lists to store the potential candidates, *i.e.*, PLs for potential local lists and PG for potential global lists, and output the answers from PG in each sliding window.

over a number of N graphs is defined in Eq.(1).

$$\phi(g, q) = \frac{NN_{gq} - N_g N_q}{\sqrt{N_g(N - N_g)N_q(N - N_q)}} \quad (1)$$

In Eq.(1), N_g , N_q , and N_{gq} denote, for the N graphs, the number of graphs containing g , q , and g and q , respectively. For clarity of presentation, when these variables are collected from the whole window D , the correlation between g and q is denoted as $\phi_D(g, q)$; when they are computed over a local batch of graphs (*e.g.*, G_j), the correlation is represented as $\phi_L(g, q)$.

3. PROPOSED METHOD

A straightforward method to solve our problem is to recompute all correlations for each sliding window, which is computationally inefficient for data streams. In this paper, we utilize Hoeffding bound [2] to discover a set of candidates in each batch, and employ a *global-local check* scheme with two levels of candidate lists to maintain these potential graphs.

3.1 Hoeffding Bound for Candidate Generation

Suppose there is a sequence of independently draw samples $\{\phi_1, \phi_2, \dots, \phi_n\}$, each within in range $R = [a, b]$. Hoeffding bound [2] gives certain probabilistic guarantees on the estimation of statistics about the underlying data. More specifically, let r be the expectation of these points, and the estimated mean $\bar{r} = \frac{1}{n} \sum_{i=1}^n \phi_i$. Hoeffding bound states, for any ϵ ,

$$\Pr\{|r - \bar{r}| \geq \epsilon\} \leq 2e^{-\frac{2n\epsilon^2}{R^2}} \quad (2)$$

Let the right side of Eq.(2) be $\delta = 2e^{-\frac{2n\epsilon^2}{R^2}}$, Hoeffding bound indicates that with probability $\leq \delta$, the estimated mean \bar{r} is beyond $\pm\epsilon$ of r ; or in other words, with confidence $1 - \delta$, the estimated mean \bar{r} is within ϵ of the expectation. From above equation, we have

$$\epsilon = \sqrt{\frac{R^2 \ln(2/\delta)}{2n}} \quad (3)$$

Algorithm 1: Hoe-PGPL(G_i, PG, PLs, k, w, m)

```
1  $S_q^{G_i} = \bigcup \{g|q \subseteq g, g \in G_i\}, N_q = |S_q^{G_i}|;$ 
2 Get top- $k$  correlated graphs in batch  $G_i$  from  $S_q^{G_i}$ ;
3 Get the  $k^{th}$  correlation value  $\phi_L$  and error  $\epsilon_s$ ;
4  $PL_i \leftarrow \bigcup \{g|\phi_L(g, q) \geq \phi_L - 2\epsilon_s, g \subseteq g', g' \in G_i\};$ 
5 foreach  $g \in PG$  do
6   Increasing  $N_g, N_{gq}$ , and  $N_q$ ;
7   if some batch is deleted then
8     Decrease  $N_g, N_{gq}$ , and  $N_q$ ;
9 Merge( $PL_i, PG, PLs$ );
10 Re-compute the top- $k$  correlated graphs based on
    PG, adding them to  $A_g$ ;
11 if  $|PG| > m * k$  then
12   Prune PG;
13 return  $A_g$ ;
```

In our problem settings, we assume the correlation over data stream is a random variable, and stream data arrive batch by batch. In each batch we estimate $\bar{r} = \phi_L(g, q)$, whose expected correlation value in a sliding window (denoted by $\phi_D(g, q)$) is at least $\phi_L(g, q) - \epsilon$ with high probability $1 - \delta$. Specifically, the $\phi_L(g, q)$ is computed in a batch over a number of $|G_i|$ graphs, and we assume this computed result can be viewed as an average of $\phi_L(g, q)$.

Based on the Hoeffding bound, we mine a set of potential top- k correlated graphs in each batch over stream. For each candidate g , we have $\phi_L(g, q) > \phi_L - 2\epsilon_s$. Here, $\epsilon_s = \sqrt{\frac{R^2 \ln(2/\delta)}{2|G_i|}}$, and ϕ_L is the k^{th} smallest correlation value with q (all candidates are sorted in a descending order based on their correlation values), which can be discovered by using the method in [5] or [3].

3.2 Framework of Hoe-PGPL Algorithm

After discovering the potential k -correlated graphs in each batch, we use two levels of candidate lists to maintain these candidates, *i.e.*, a potential global list (PG) and a set of potential local lists (PLs). PG and PLs are also shown in Fig. 1. The detailed procedures of the proposed Hoe-PGPL algorithm is shown in Algorithm 1. In addition to the PG, PLs, and G_i , Hoe-PGPL takes several parameters as inputs. Specifically, k indicates the number of answers (graphs) to be returned; w represents the size of the sliding window (in terms of number of batches); and m is a parameter controlling the maximum number of candidates in PG. The output of Hoe-PGPL is the top- k correlated graphs A_g in the current sliding window.

Hoe-PGPL in Algorithm 1 mainly consists of four parts: (1) Hoeffding bound based candidate generation (steps 1-4); (2) Increasing and decreasing frequency information in PG (steps 5-8); When new batches arrive or become outdated, for each $g \in PG$, we update the frequency information in the window D . (3) Inserting new candidates into PG (steps 9), which is a key step for estimating the correlations of candidates (the details are reported in Algorithm 2); (4) Pruning PG (steps 11-12). If $|PG|$ exceeds $m * k$, we delete some unpromising candidates. Let $\epsilon_w = \sqrt{\frac{R^2 \ln(2/\delta)}{2w|G_i|}}$ and ϕ_D be the k -th correlated graph in PG, we delete candidates whose correlation is smaller than $\phi_D - 2\epsilon_w$.

Algorithm 2: Merge(PL_i, PG, PLs)

```
1  $\tau_{min} = \arg \min_{g_0 \in PG} \phi_D(g_0, q);$ 
2 foreach  $\tau \in PL_i$  and  $\tau \notin PG$  do
3   foreach  $PL_j, j \in [i - w + 1, i - 1]$  do
4     if  $\tau \in PL_j$  then
5        $g' = \tau$ ;
6     else
7        $g' = \arg \min_{g \in PL_j} \phi_L(g, q);$ 
8   Add the statistics of  $g'$  in batch  $G_j$  into  $N_\tau, N_{\tau q}$ ,
    and  $N_q$  for the window;
9   if  $\phi_D(\tau_{min}, q) < \phi_D(\tau, q)$  then
10     Further revise the statistics of  $\tau$ ;
11 Insert  $\tau$  to PG;
```

3.3 Global-local Check Scheme

In graph stream mining, because we cannot afford to record all information about the stream data, the complications normally arise from the situations a pattern τ , which is insignificant in previous observations, may emerge and become significant. Because τ is insignificant in previous observations, its information in the historical stream data is not recorded and therefore lost. We regard such patterns as **emerging candidate patterns**. Because we are not allowed to check the historical graph data to find τ 's frequency in the previous batches of the window, the correlation value of $\phi_D(\tau, q)$ will have to be estimated, which may, in turn, introduce errors. We use two level of candidate list to handle this problem.

- **Potential Global List (PG):** A global candidate list PG is used to store the global frequency information of each candidate (N_g, N_{gq} , and N_q) in the sliding window D , from which candidates are sorted according to their correlation values.
- **Potential Local Lists (PLs):** We also use a set of PLs to record the frequency information of candidates (N_g, N_{gq} , and N_q) in a local batch. Specifically, we use a list $PL_j, j \in [i - w + 1, i]$ to store the candidates retrieved from batch G_j .

Basically, there are two rules to estimate each emerging candidate pattern τ 's frequency. (1) From a global perspective, its estimated value $\hat{\phi}_D(\tau, q)$ should be less than the minimum correlation value ($\phi_D(\tau_{min}, q)$) in PG if it is not stored in PG. (2) From a local perspective, its estimated value $\hat{\phi}_L(\tau, q)$ in each batch should be less than the minimum correlation value ($\phi_L(\tau_{min}, q)$) in a local batch if it is not stored in a local list PL_j .

Our scheme for maintaining PG and PLs is illustrated in Algorithm 2. Firstly we go through PL_{i-w+1} to PL_{i-1} to collect the estimated frequency information of each emerging candidate τ in a local perspective (steps 3-8 in Algorithm 2), which assures that $\phi_L(\tau, q)$ in $PL_j, j \in [i - w + 1, i - 1]$ is close to its true value in batch G_j . Step 8 uses the frequency information of g to estimate candidate τ in batch G_j . After collecting the frequency information of τ from PLs, we check it in PG from a global perspective (steps 9-10). Let

τ_{min} be the graph with minimum correlation value in PG, if $\phi_D(\tau_{min}, q)$ is smaller than the updated correlation value $\phi_D(\tau, q)$, we further revise the frequency information of g as we want $\phi_D(\tau, q) < \phi_D(\tau_{min}, q)$ in a global perspective. In our implementation, if $\tau \notin PL_j, j \in [i - w + 1, i - 1]$, we replace the frequency information of τ in batch G_j with the frequency information of τ_{min} in that batch. By doing so, the newly emerging candidates can be inserted into the PG with a precise estimation of its genuine correlation with query graph q . As a result, the ranking of PG will be more accurate and the query results can be more precise.

4. EXPERIMENTAL RESULTS

To validate the performance of our method, we collect graph data stream from a real-world NCI graph database¹, which contains compound structures of cancer and AIDS data. The original dataset contains about 249,000 graphs. After preprocessing and removing some disconnected graphs, we have a data stream with about 233,000 graphs.

The query graphs in our experiments consists of 50 graphs, each of which is randomly selected with a support in the whole data stream is less than 0.05. In addition to the system runtime and memory consumption, Precision and Recall are used to measure the results of our algorithm. Suppose A_g and T_g denotes the returned answers and true answers respectively, then $Precision = |T_g \cap A_g|/|A_g|$ and $Recall = |T_g \cap A_g|/|T_g|$. An algorithm is perfectly accurate if both precision and recall are equal to 1.

We compare our algorithms with an exhaustive search method (denoted by **Fixwin**) in terms of system runtime and memory consumption. When implementing Fixwin, *i.e.*, whenever a new batch of graphs arrives, Fixwin restarts to mine the top- k correlated graphs in the sliding window from scratch. To justify our estimated scheme, we compare Hoe-PGPL with a **Hoe-PG** scheme, which only uses PG to estimate the emerging candidate pattern τ (PLs are not employed). Unless specified otherwise, we set the default values $\delta=0.05$, $k=30$, $w=10$, $m=5$, and $|G_i|=5000$.

In Fig. 2, we report the memory consumptions *w.r.t.* different top- k values. The result shows that the exhaustive search method (Fixwin) consumes several times more memory comparing to the Hoe-PGPL, mainly because Fixwin needs to store the graph data in the entire window while Hoe-PGPL only saves the most up-to-date batch of data and some candidates in PG and PLs. Meanwhile, the runtime performance in Fig. 2 demonstrates that Hoeffding bound based methods are about 5-10 times faster than the Fixwin method. On the other hand, when increasing the top- k values, the results in Fig. 3 show that the precision and the recall values of Hoe-PG and Hoe-PGPL both decline, with Hoe-PGPL showing much better performance than Hoe-PG. This demonstrates that by combining the two levels of candidate lists and the dynamic estimation of missing values of emerging candidate patterns (τ) in the historical batches, our method is able to maintain high effectiveness and high accuracy for stream-based top- k correlated graph query.

5. CONCLUSION

In this paper, we studied top- k correlated graph query from data stream, which discovers a query graph q 's top- k correlated graphs in a sliding window covering a portion of

¹<http://cactus.nci.nih.gov/ncidb2/download.html>

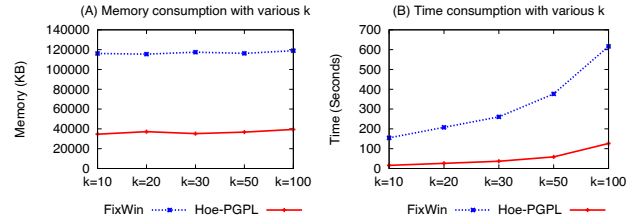


Figure 2: Memory and system runtime consumption VS. k values.

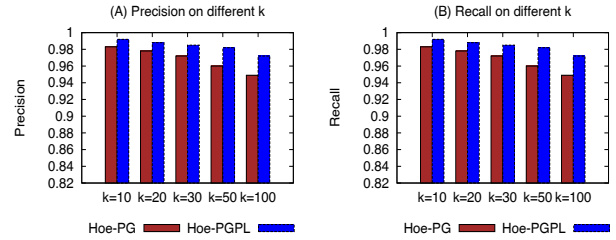


Figure 3: Precision and Recall VS. k values.

stream data records. We argued that, for data streams with dynamic increasing volumes, simple exhaustive search for top- k correlated graphs needs to store the graph data within the sliding window and repeatedly carry out the query process, which is computationally expensive and memory intensive. By employing Hoeffding bound to discover a set of candidates in each batch and utilizing a global-local check scheme that integrates both potential global candidates (PG) and local candidates (PLs), we can accurately estimate the correlation of each candidates in PG. Experimental results demonstrate that Hoe-PGPL is several times more efficient than the exhaustive search method with respect to time and memory consumptions, and achieves high performance in terms of query precision and recall.

6. ACKNOWLEDGMENTS

This work was supported by the Australian Research Council (ARC) Future Fellowship under Grant No. FT100100971.

7. REFERENCES

- [1] S. A. Cook. The complexity of theorem-proving procedures. In *ACM Symposium on Theory of Computing*, pages 151–158, 1971.
- [2] W. Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58:13–30, 1963.
- [3] Y. Ke, J. Cheng, and N. Wilfred. Correlation search in graph databases. In *KDD*, pages 390–399, 2007.
- [4] Y. Ke, J. Cheng, and N. Wilfred. Correlated pattern mining in quantitative databases. *ACM Trans. on Database Systems*, 33:1–45, 2008.
- [5] Y. Ke, J. Cheng, and J. X. Yu. Top- k Correlative Graph Mining. In *SIAM International Conf. on Data Mining*, pages 1038–1049, 2009.
- [6] W. Zhou and H. Xiong. Volatile correlation computation: a checkpoint view. In *KDD*, pages 848–856, 2008.

[CIKM12](#)**The 21st ACM International Conference on Information and Knowledge Management**

Oct. 29 - Nov. 2, 2012, Maui, Hawaii, USA

Reviews For Paper**Track** Posters**Paper ID** 1780**Title** Continuous Top-k Correlated Graph Query for Data Streams**Masked Reviewer ID:** Assigned_Reviewer_1**Review:**

Question	
Overall Recommendation	Weak Reject: I don't really like the paper but won't object strongly against acceptance
Detailed Comments	<p>In the paper, the authors propose Hoe-PGGL to identify top-k correlated graphs from data stream. Experiments have been conducted to verify the effectiveness and efficiency of the proposed approach.</p> <p>(1) Abstract. Our method represents the first research endeavor for data stream based top-k correlated graph query: downgrade the tune "the first research".</p> <p>(2) Section 3.2. Equation 2 is meaningless since the right hand size is always equal or larger 2.</p> <p>(3) I cannot see how equation 3 is derived from equation 2. It is like a magic - it is mostly a wrong claim.</p> <p>Equations 2 and 3 are quite important for the entire paper. I should be clarified.</p>

Masked Reviewer ID: Assigned_Reviewer_2**Review:**

Question	
Overall Recommendation	Weak Accept: I quite like the paper but won't object strongly against rejection
Detailed Comments	<p>The paper studies an existing problem on a new setting; that is, correlation search over graph streams. Specifically, given a query graph and a continuous graph stream, it is to find the top-k correlated graphs between the query and the data stream over sliding windows.</p> <p>Strong points:</p> <p>1) It investigates an existing problem in a streaming setting, which may have potential applications;</p> <p>2) It proposes to use Hoeffding bound for candidate generation, and the potential global and local list data structure are introduced to support effective estimation.</p> <p>3) Experimental result is good, compared with the baseline method. It is interesting to see the statistical bound has good performance on structure related problems.</p> <p>4) With clear presentation, the paper is easily readable.</p>

Weak points:

- 1) Formatting problem: "stream-based top-k..." in Section 4.
- 2) Grammatical mistake: "Fixwin restart to mine..." in Section 4.



- [Main Page](#)
- [CIKM'12](#)
- [BooksOnline'12](#)
- [CDMW'12](#)
- [CloudDB'12](#)
- [CrowdSens'12](#)
- [DOLAP'12](#)
- [DTMBIO'12](#)
- [DUBMMSM'12](#)
- [ESAIR'12](#)
- [IKM4DR'12](#)
- [MIXHS'12](#)
- [PIKM'12](#)
- [PLEAD'12](#)
- [SHB'12](#)
- [Web-KR'12](#)
- [WIDM'12](#)
- [Compilation Author Index](#)
- [CIKM 2013 Call for Papers](#)

CIKM 2012 Proceedings

Xuewen Chen, Guy Lebanon, Haixun Wang, Mohammed J. Zaki, & Wolfgang Gatterbauer (editors), Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM'12, October 29–November 2, 2012, Maui, HI, USA. ACM 2012, ISBN 978-1-4503-1156-4.

Table of Contents (Complete Table of Contents)

- CIKM 2012 General Chair's Welcome**
Xuewen Chen (*Wayne State University*)
- Program Committee Chairs' Welcome**
Guy Lebanon (*Georgia Institute of Technology*)
Haixun Wang (*Microsoft Research Asia*)
Mohammed J. Zaki (*Rensselaer Polytechnic Institute*)
- CIKM 2012 Conference Organization**
- CIKM 2012 DB Track Senior Program Committee**
- CIKM 2012 IR Track Senior Program Committee**
- CIKM 2012 KM Track Senior Program Committee**
- CIKM 2012 External Reviewers**
- CIKM 2012 Sponsors & Supporters**
- CIKM 2012 Author Index**

CIKM 2012 Online Program at a Glance

- | | |
|---|--|
| Keynote Addresses | Session 21 -- KM Track: Spatial and Temporal Methods |
| Session 1 -- KM Track: Recommender Systems | Session 22 -- IR Track: Web Search |
| Session 2-- KM Track: Pattern Mining | Session 23 -- DB Track: Web Data Management |
| Session 3 -- IR Track: Evaluation Methodologies | Session 24 -- KM Track: Information Extraction |
| Session 4 -- IR Track: Social Media Search | Session 25 -- IR Track: Topic Modeling and Content and Sentiment analysis |
| Session 5 -- KM Track: Link and Graph Mining | Session 26 -- DB Track: Query Processing, Optimization and Performance |
| Session 6 -- IR Track: Language Technologies | Session 27 -- KM Track: Classification and Semantic Methods |
| Session 7 -- DB Track: Graph and Knowledge Base | Session 28 -- IR Track: Multimedia and User Feedback |
| Session 8 -- DB Track: Temporal, Spatial and Multimedia Databases | Session 29 -- DB Track: Emerging and Advanced Topics |
| Session 9 -- KM Track: Matrix Methods and Anomaly Detection | Session 30 -- KM Track: Novel Applications |
| Session 10 -- KM Track: Social Networks | Session 31 -- IR Track: Social Networks |
| Session 11 -- IR Track: Advertising | Knowledge Management Short Paper Session |
| Session 12 -- IR Track: System Architecture, Distributed IR, Scalability | Information Retrieval Short Paper Session |
| Session 13 -- KM Track: Advertisement and Products | Databases Short Paper Session |

Session 14 -- KM Track: Clustering	Knowledge Management Poster Session
Session 15 -- IR Track: Recommendation Systems	Information Retrieval Poster Session
Session 16 -- IR Track: Digital Libraries and Citation Analysis	Databases Poster Session
Session 17 -- KM Track: Text Mining	Knowledge Management Demonstration Session
Session 18 -- IR Track: Formal Retrieval Models and Learning to Rank	Information Retrieval Demonstration Session
Session 19 -- DB Track: Probabilistic and Uncertain Data	Databases Demonstration Session
Session 20 -- DB Track: Top-k and Nearest Neighbor Queries	Workshop Summaries

Keynote Addresses

User Engagement: The Network Effect Matters! (Page 1)

Ricardo Baeza-Yates (*Yahoo! Labs*)
(joint work with Mounia Lalmas, *Yahoo! Labs*)

Learning Similarity Measures Based on Random Walks (Page 3)

William W. Cohen (*Carnegie Mellon University*)

Compressed Data Structures with Relevance (Page 4)

Jeffrey Scott Vitter (*The University of Kansas*)

Session 1 - KM Track: Recommender Systems

LogUCB: An Explore-Exploit Algorithm for Comments Recommendation (Page 6)

Dhruv Kumar Mahajan (*Microsoft Research Asia*)
Rajeev Rastogi (*Amazon*)
Charu Tiwari (*Yahoo! Labs*)
Adway Mitra (*IISC*)

DQR: A Probabilistic Approach to Diversified Query Recommendation (Page 16)

Ruirui Li (*The University of Hong Kong*)
Ben Kao (*The University of Hong Kong*)
Bin Bi (*University of California, Los Angeles*)
Reynold Cheng (*The University of Hong Kong*)
Eric Lo (*Hong Kong Polytechnic University*)

Dynamic Covering for Recommendation Systems (Page 26)

Ioannis Antonellis (*oDesk*)
Anish Das Sarma (*Google Research*)
Shaddin Dughmi (*Microsoft Research*)

MEET: A Generalized Framework for Reciprocal Recommender Systems (Page 35)

Lei Li (*Florida International University*)
Tao Li (*Florida International University*)

Social Contextual Recommendation (Page 45)

Meng Jiang (*Beijing Key Laboratory & Tsinghua University*)
Peng Cui (*Beijing Key Laboratory & Tsinghua University*)
Rui Liu (*Tsinghua University*)
Qiang Yang (*Hong Kong University of Science and Technology*)
Fei Wang (*IBM T.J. Watson Research Center*)
Wenwu Zhu (*Beijing Key Laboratory & Tsinghua University*)
Shiqiang Yang (*Beijing Key Laboratory & Tsinghua University*)

Session 2 - KM Track: Pattern Mining

Mining High Utility Itemsets without Candidate Generation (Page 55)

Mengchi Liu (*Wuhan University & Carleton University*)
Junfeng Qu (*Wuhan University*)

A General Framework to Encode Heterogeneous Information Sources for Contextual Pattern Mining (Page 65)

Weishan Dong (*IBM Research - China*)
Wei Fan (*Huawei Noah's Ark Lab*)
Lei Shi (*Chinese Academy of Sciences*)
Changjin Zhou (*IBM Research - China*)
Xifeng Yan (*University of California, Santa Barbara*)

Incorporating Occupancy into Frequent Pattern Mining for High Quality Pattern Recommendation (Page 75)

Linpeng Tang (*Shanghai Jiao Tong University & HP Labs China*)
Lei Zhang (*University of Science and Technology of China & HP Labs China*)
Ping Luo (*HP Labs China*)

Min Wang (*HP Labs China*)

PARMA: A Parallel Randomized Algorithm for Approximate Association Rules Mining in MapReduce (Page 85)

Matteo Riondato (*Brown University*)

Justin A. DeBrabant (*Brown University*)

Rodrigo Fonseca (*Brown University*)

Eli Upfal (*Brown University*)

Interactive Pattern Mining on Hidden Data: A Sampling-Based Solution (Page 95)

Mansurul Bhuiyan (*Indiana University - Purdue University, Indianapolis*)

Snehasis Mukhopadhyay (*Indiana University - Purdue University, Indianapolis*)

Mohammad Al Hasan (*Indiana University - Purdue University, Indianapolis*)

Session 3 - IR Track: Evaluation Methodologies

An Analysis of Systematic Judging Errors in Information Retrieval (Page 105)

Gabriella Kazai (*Microsoft Research*)

Nick Craswell (*Microsoft*)

Emine Yilmaz (*Microsoft Research*)

S.M.M. Tahaghoghi (*Microsoft*)

On Caption Bias in Interleaving Experiments (Page 115)

Katja Hofmann (*University of Amsterdam*)

Fritz Behr (*Microsoft*)

Filip Radlinski (*Microsoft*)

Alternative Assessor Disagreement and Retrieval Depth (Page 125)

William Webber (*University of Maryland*)

Praveen Chandar (*University of Delaware*)

Ben Carterette (*University of Delaware*)

Incorporating Variability in User Behavior into Systems Based Evaluation (Page 135)

Ben Carterette (*University of Delaware*)

Evangelos Kanoulas (*University of Sheffield*)

Emine Yilmaz (*Microsoft Research*)

Constructing Test Collections by Inferring Document Relevance via Extracted Relevant Information (Page 145)

Shahzad Rajput (*Northeastern University*)

Matthew Ekstrand-Abueg (*Northeastern University*)

Virgil Pavlu (*Northeastern University*)

Javed A. Aslam (*Northeastern University*)

Session 4 - IR Track: Social Media Search

Twevent: Segment-based Event Detection from Tweets (Page 155)

Chenliang Li (*Nanyang Technological University*)

Aixin Sun (*Nanyang Technological University*)

Anwitaman Datta (*Nanyang Technological University*)

Making Your Interests Follow You on Twitter (Page 165)

Marco Pennacchiotti (*eBay Inc.*)

Fabrizio Silvestri (*ISTI - CNR*)

Hossein Vahabi (*ISTI - CNR & IMT Lucca*)

Rossano Venturini (*University of Pisa*)

Generating Event Storylines from Microblogs (Page 175)

Chen Lin (*Xiamen University*)

Chun Lin (*Xiamen University*)

Jingxuan Li (*Florida International University*)

Dingding Wang (*Florida International University*)

Yang Chen (*Duke University*)

Tao Li (*Florida International University*)

Social Book Search: Comparing Topical Relevance Judgements and Book Suggestions for Evaluation (Page 185)

Marijn Koolen (*University of Amsterdam*)

Jaap Kamps (*University of Amsterdam*)

Gabriella Kazai (*Microsoft Research*)

Content-Based Crowd Retrieval on the Real-Time Web (Page 195)

Krishna Y. Kamath (*Texas A&M University*)

James Caverlee (*Texas A&M University*)

Session 5 - KM Track: Link and Graph Mining

Graph Classification: A Diversified Discriminative Feature Selection Approach (Page 205)

Yuanyuan Zhu (*The Chinese University of Hong Kong*)

Jeffrey Xu Yu (*The Chinese University of Hong Kong*)

Hong Cheng (*The Chinese University of Hong Kong*)

Lu Qin (*The Chinese University of Hong Kong*)

Multi-Scale Link Prediction (Page 215)

Donghuk Shin (*University of Texas at Austin*)

(Return to Top)

(Return to Top)

Si Si (*University of Texas at Austin*)

Inderjit S. Dhillon (*University of Texas at Austin*)

An Analysis of How Ensembles of Collective Classifiers Improve Predictions in Graphs (Page 225)

Hoda Eldardiry (*Palo Alto Research Center*)

Jennifer Neville (*Purdue University*)

Density Index and Proximity Search in Large Graphs (Page 235)

Nan Li (*University of California, Santa Barbara*)

Xifeng Yan (*University of California, Santa Barbara*)

Zhen Wen (*IBM Research*)

Arijit Khan (*University of California, Santa Barbara*)

Gelling, and Melting, Large Graphs by Edge Manipulation (Page 245)

Hanghang Tong (*IBM T.J. Watson Research Center*)

B. Aditya Prakash (*Virginia Tech.*)

Tina Eliassi-Rad (*Rutgers University*)

Michalis Faloutsos (*University of California, Riverside*)

Christos Faloutsos (*Carnegie Mellon University*)

Session 6 - IR Track: Language Technologies

One Seed to Find Them All: Mining Opinion Features via Association (Page 255)

Zhen Hai (*Nanyang Technological University*)

Kuiyu Chang (*Nanyang Technological University*)

Gao Cong (*Nanyang Technological University*)

Topic-Driven Reader Comments Summarization (Page 265)

Zongyang Ma (*Nanyang Technological University*)

Aixin Sun (*Nanyang Technological University*)

Quan Yuan (*Nanyang Technological University*)

Gao Cong (*Nanyang Technological University*)

Visualizing Timelines: Evolutionary Summarization via Iterative Reinforcement between Text and Image Streams (Page 275)

Rui Yan (*Peking University & National Taiwan University*)

Xiaojun Wan (*Peking University*)

Mirella Lapata (*University of Edinburgh*)

Wayne Xin Zhao (*Peking University*)

Pu-Jen Cheng (*National Taiwan University*)

Xiaoming Li (*Beihang University & Peking University*)

Fast Multi-Task Learning for Query Spelling Correction (Page 285)

Xu Sun (*Cornell University*)

Anshumali Shrivastava (*Cornell University*)

Ping Li (*Cornell University*)

Cross-Argument Inference for Implicit Discourse Relation Recognition (Page 295)

Yu Hong (*Soochow University*)

Xiaopei Zhou (*Soochow University*)

Tingting Che (*Soochow University*)

Jianmin Yao (*Soochow University*)

Guodong Zhou (*Soochow University*)

Qiaoming Zhu (*Soochow University*)

Session 7 - DB Track: Graph and Knowledge Base

Interpreting Keyword Queries over Web Knowledge Bases (Page 305)

Jeffrey Pound (*University of Waterloo*)

Alexander K. Hudek (*University of Waterloo*)

Ihab F. Ilyas (*Qatar Computing Research Institute*)

Grant Weddell (*University of Waterloo*)

RDF Pattern Matching using Sortable Views (Page 315)

Zhihong Chong (*Southeast University*)

He Chen (*Southeast University*)

Zhenjie Zhang (*Illinois at Singapore Pte. Ltd.*)

Hu Shu (*Southeast University*)

Guilin Qi (*Southeast University*)

Aoying Zhou (*East China Normal University*)

Efficient Algorithms for Generalized Subgraph Query Processing (Page 325)

Wenqing Lin (*Nanyang Technological University*)

Xiaokui Xiao (*Nanyang Technological University*)

James Cheng (*Nanyang Technological University*)

Sourav S. Bhowmick (*Nanyang Technological University*)

G-SPARQL: A Hybrid Engine for Querying Large Attributed Graphs (Page 335)

Sherif Sakr (*NICTA and University of New South Wales*)

Sameh Elnikety (*Microsoft Research*)

Yuxiong He (*Microsoft Research*)

A Graph-Based Approach for Ontology Population with Named Entities (Page 345)Wei Shen (*Tsinghua University*)Jianyong Wang (*Tsinghua University*)Ping Luo (*HP Labs China*)Min Wang (*HP Labs China*)**Session 8 - DB Track: Temporal, Spatial and Multimedia Databases****Decomposition-by-Normalization (DBN): Leveraging Approximate Functional Dependencies for Efficient Tensor****Decomposition** (Page 355)Mijung Kim (*Arizona State University*)K. Selçuk Candan (*Arizona State University*)**A Filter-based Protocol for Continuous Queries over Imprecise Location Data** (Page 365)Yifan Jin (*University of Hong Kong*)Reynold Cheng (*University of Hong Kong*)Ben Kao (*University of Hong Kong*)Kam-Yiu Lam (*City University of Hong Kong*)Yinuo Zhang (*University of Southern California*)**Leveraging Read Rates of Passive RFID Tags for Real-Time Indoor Location Tracking** (Page 375)Da Yan (*The Hong Kong University of Science and Technology*)Zhou Zhao (*The Hong Kong University of Science and Technology*)Wilfred Ng (*The Hong Kong University of Science and Technology*)**Location-Aware Instant Search** (Page 385)Ruicheng Zhong (*Tsinghua University*)Ju Fan (*Tsinghua University*)Guoliang Li (*Tsinghua University*)Kian-Lee Tan (*National University of Singapore*)Lizhu Zhou (*Tsinghua University*)**Indexing Uncertain Spatio-Temporal Data** (Page 395)Tobias Emrich (*Ludwig-Maximilians-Universität-München*)Hans-Peter Kriegel (*Ludwig-Maximilians-Universität-München*)Nikos Mamoulis (*University of Hong Kong*)Matthias Renz (*Ludwig-Maximilians-Universität-München*)Andreas Züfle (*Ludwig-Maximilians-Universität-München*)**Session 9 - KM Track: Matrix Methods and Anomaly Detection****Local Anomaly Descriptor: A Robust Unsupervised Algorithm for Anomaly Detection Based on Diffusion Space** (Page 405)Hao Huang (*Stony Brook University*)Hong Qin (*Stony Brook University*)Shinjae Yoo (*Brookhaven National Laboratory*)Dantong Yu (*Brookhaven National Laboratory*)**Fast and Reliable Anomaly Detection in Categorical Data** (Page 415)Leman Akoglu (*Carnegie Mellon University*)Hanghang Tong (*IBM T.J. Watson Research Center*)Jilles Vreeken (*University of Antwerp*)Christos Faloutsos (*Carnegie Mellon University*)**TALMUD - Transfer Learning for Multiple Domains** (Page 425)Orly Moreno (*Ben-Gurion University of the Negev & Deutsche Telekom Laboratories*)Bracha Shapira (*Ben-Gurion University of the Negev & Deutsche Telekom Laboratories*)Lior Rokach (*Ben-Gurion University of the Negev & Deutsche Telekom Laboratories*)Guy Shani (*Ben-Gurion University of the Negev & Deutsche Telekom Laboratories*)**Utilizing Common Substructures to Speedup Tensor Factorization for Mining Dynamic Graphs** (Page 435)Wei Liu (*The University of Melbourne*)Jeffrey Chan (*The University of Melbourne*)James Bailey (*The University of Melbourne*)Christopher Leckie (*The University of Melbourne*)Ramamohanarao Kotagiri (*The University of Melbourne*)**Session 10 - KM Track: Social Networks****Predicting Emerging Social Conventions in Online Social Networks** (Page 445)Farshad Kooti (*MPI-SWS*)Winter A. Mason (*Stevens Institute of Technology*)Krishna P. Gummadi (*MPI-SWS*)Meeyoung Cha (*Korea Advanced Institute of Science and Technology*)**Collective Intelligence in the Online Social Network of Yahoo!Answers and Its Implications** (Page 455)Ze Li (*Clemson University*)Haiying Shen (*Clemson University*)Joseph Edward Grant (*Clemson University*)**From Face-to-Face Gathering to Social Structure** (Page 465)[\(Return to Top\)](#)[\(Return to Top\)](#)[\(Return to Top\)](#)

[From Face-to-Face Gathering to Social Structure](#) (Page 466)

Chunyan Wang (*Stanford University*)

Mao Ye (*HP Labs*)

Wang-Chien Lee (*The Pennsylvania State University*)

Delineating Social Network Data Anonymization via Random Edge Perturbation (Page 475)

Mingqiang Xue (*Institute for Infocomm Research*)

Panagiotis Karras (*Rutgers University*)

Chedy Raïssi (*INRIA Nancy*)

Panos Kalnis (*KAUST*)

Hung Keng Pung (*National University of Singapore*)

Session 11 - IR Track: Advertising

Multiview Hierarchical Bayesian Regression Model and Application to Online Advertising (Page 485)

Tianbing Xu (*University of California, Irvine*)

Ruofei Zhang (*Yahoo! Labs*)

Zhen Guo (*Yahoo! Labs*)

Visual Appearance of Display Ads and Its Effect on Click Through Rate (Page 495)

Javad Azimi (*Oregon State University*)

Ruofei Zhang (*Yahoo! Labs*)

Yang Zhou (*Yahoo! Labs*)

Vidhya Navalpakkam (*Google*)

Jianchang Mao (*Yahoo! Labs*)

Xiaoli Fern (*Oregon State University*)

The Wisdom of Advertisers: Mining Subgoals via Query Clustering (Page 505)

Takehiro Yamamoto (*Microsoft Research Asia, Kyoto University & JST Crest*)

Tetsuya Sakai (*Microsoft Research Asia*)

Mayu Iwata (*Microsoft Research Asia & Osaka University*)

Chen Yu (*Microsoft Research Asia*)

Ji-Rong Wen (*Microsoft Research Asia*)

Katsumi Tanaka (*Kyoto University*)

Sequential Selection of Correlated Ads by POMDPs (Page 515)

Shuai Yuan (*University College London*)

Jun Wang (*University College London*)

Session 12 - IR Track: System Architecture, Distributed IR, Scalability

Diversity in Blog Feed Retrieval (Page 525)

Mostafa Keikha (*University of Lugano*)

Fabio Crestani (*University of Lugano*)

W. Bruce Croft (*University of Massachusetts, Amherst*)

Efficient Retrieval of Recommendations in a Matrix Factorization Framework (Page 535)

Noam Koenigstein (*Tel-Aviv University*)

Parikshit Ram (*Georgia Institute of Technology*)

Yuval Shavitt (*Tel-Aviv University*)

KORE: Keyphrase Overlap Relatedness for Entity Disambiguation (Page 545)

Johannes Hoffart (*Max Planck Institute for Informatics*)

Stephan Seufert (*Max Planck Institute for Informatics*)

Dat Ba Nguyen (*Max Planck Institute for Informatics*)

Martin Theobald (*Max Planck Institute for Informatics*)

Gerhard Weikum (*Max Planck Institute for Informatics*)

Shard Ranking and Cutoff Estimation for Topically Partitioned Collections (Page 555)

Anagha Kulkarni (*Carnegie Mellon University*)

Almer S. Tigelaar (*University of Twente*)

Djoerd Hiemstra (*University of Twente*)

Jamie Callan (*Carnegie Mellon University*)

Session 13 - KM Track: Advertisement and Products

Daily-Deal Selection for Revenue Maximization (Page 565)

Theodoros Lappas (*Boston University*)

Evimaria Terzi (*Boston University*)

Enabling Direct Interest-Aware Audience Selection (Page 575)

Ariel Fuxman (*Microsoft Research*)

Anitha Kannan (*Microsoft Research*)

Zhenhui Li (*University of Illinois*)

Panayiotis Tsaparas (*University of Ioannina*)

Influence Propagation in Adversarial Setting: How to Defeat Competition with Least Amount of Investment (Page 585)

Shahrad Shirazipourazad (*Arizona State University*)

Brian Bogard (*Arizona State University*)

Harsh Vachhani (*Arizona State University*)

Arunabha Sen (*Arizona State University*)

Paul Horn (*Harvard University*)

(Return to Top)

(Return to Top)

(Return to Top)

Large-scale Item Categorization for e-Commerce (Page 595)Dan Shen (*dianping.com*)Jean-David Ruvini (*eBay Research Labs*)Badrul Sarwar (*eBay Research Labs*)**Matching Product Titles using Web-based Enrichment** (Page 605)Vishrawas Gopalakrishnan (*Suny Buffalo*)Suresh Parthasarathy Iyengar (*Yahoo! Labs*)Amit Madaan (*TheFind, Inc.*)Rajeev Rastogi (*Amazon*)Srinivasan Sengamedu (*Komli Labs*)**Session 14 - KM Track: Clustering****Scalable Clustering of Signed Networks Using Balance Normalized Cut** (Page 615)Kai-Yang Chiang (*University of Texas at Austin*)Joyce Jiyoung Whang (*University of Texas at Austin*)Inderjit S. Dhillon (*University of Texas at Austin*)**Maximum Margin Clustering on Evolutionary Data** (Page 625)Xuhui Fan (*University of Technology, Sydney*)Longbing Cao (*University of Technology, Sydney*)Xia Cui (*University of Technology, Sydney*)Lin Zhu (*Shanghai Jiao Tong University*)Yew-Soon Ong (*Nanyang Technological University*)**Document-Topic Hierarchies from Document Graphs** (Page 635)Tim Weninger (*University of Illinois at Urbana-Champaign*)Yonatan Bisk (*University of Illinois at Urbana-Champaign*)Jiawei Han (*University of Illinois at Urbana-Champaign*)**Improving Document Clustering Using Automated Machine Translation** (Page 645)Xiang Wang (*University of California, Davis*)Buyue Qian (*University of California, Davis*)Ian Davidson (*University of California, Davis*)**Right-Protected Data Publishing with Hierarchical Clustering Preservation** (Page 654)Michail Vlachos (*IBM Research - Zurich*)Aleksander Wiczorek (*Poznan University*)Johannes Schneider (*IBM Research - Zurich*)**Session 15 - IR Track: Recommendation Systems****Metaphor: A System for Related Search Recommendations** (Page 664)Azarias Reda (*University of Michigan*)Yubin Park (*University of Texas at Austin*)Mitul Tiwari (*LinkedIn*)Christian Posse (*LinkedIn*)Sam Shah (*LinkedIn*)**Exploring Personal Impact for Group Recommendation** (Page 674)Xingjie Liu (*The Pennsylvania State University*)Yuan Tian (*The Pennsylvania State University*)Mao Ye (*The Pennsylvania State University*)Wang-Chien Lee (*The Pennsylvania State University*)**The Efficient Imputation Method for Neighborhood-based Collaborative Filtering** (Page 684)Yongli Ren (*Deakin University*)Gang Li (*Deakin University*)Jun Zhang (*Deakin University*)Wanlei Zhou (*Deakin University*)**Multi-Faceted Ranking of News Articles using Post-Read Actions** (Page 694)Deepak Agarwal (*LinkedIn*)Bee-Chung Chen (*LinkedIn*)Xuanhui Wang (*Facebook*)**A Decentralized Recommender System for Effective Web Credibility Assessment** (Page 704)Thanasis G. Papaioannou (*École Polytechnique Fédérale de Lausanne*)Jean-Eudes Ranvier (*École Polytechnique Fédérale de Lausanne*)Alexandra Olteanu (*École Polytechnique Fédérale de Lausanne*)Karl Aberer (*École Polytechnique Fédérale de Lausanne*)**Session 16 - IR Track: Digital Libraries and Citation Analysis****Towards an Effective and Unbiased Ranking of Scientific Literature Through Mutual Reinforcement** (Page 714)Xiaorui Jiang (*Chinese Academy of Sciences*)Xiaoping Sun (*Chinese Academy of Sciences*)Hai Zhuge (*Chinese Academy of Sciences*)**A Math-Aware Search Engine for Math Question Answering System** (Page 724)Tam T. Nguyen (*Nanyang Technological University*)[\(Return to Top\)](#)[\(Return to Top\)](#)

Kuiyu Chang (*Nanyang Technological University*)
 Siu Cheung Hui (*Nanyang Technological University*)

Contextualization Using Hyperlinks and Internal Hierarchical Structure of Wikipedia Documents (Page 734)

Muhammad Ali Norozi (*Norwegian University of Science and Technology*)
 Paavo Arvola (*University of Tampere*)
 Arjen P. de Vries (*Centrum Wiskunde & Informatica*)

Understanding Book Search Behavior on the Web (Page 744)

Jin Young Kim (*University of Massachusetts*)
 Henry Feild (*University of Massachusetts*)
 Marc-Allen Cartright (*University of Massachusetts*)

Temporal Corpus Summarization Using Submodular Word Coverage (Page 754)

Ruben Sipos (*Cornell University*)
 Adith Swaminathan (*Cornell University*)
 Pannaga Shivaswamy (*Cornell University*)
 Thorsten Joachims (*Cornell University*)

Session 17 - KM Track: Text Mining

TCSST: Transfer Classification of Short & Sparse Text Using External Data (Page 764)

Guodong Long (*University of Technology, Sydney*)
 Ling Chen (*University of Technology, Sydney*)
 Xingquan Zhu (*University of Technology, Sydney*)
 Chengqi Zhang (*University of Technology, Sydney*)

The Generalized Dirichlet Distribution in Enhanced Topic Detection (Page 773)

Karla Caballero (*University of California, Santa Cruz*)
 Joel Barajas (*University of California, Santa Cruz*)
 Ram Akella (*University of California, Santa Cruz*)

Modeling Topic Hierarchies with the Recursive Chinese Restaurant Process (Page 783)

Joon Hee Kim (*Korea Advanced Institute of Science and Technology*)
 Dongwoo Kim (*Korea Advanced Institute of Science and Technology*)
 Suin Kim (*Korea Advanced Institute of Science and Technology*)
 Alice Oh (*Korea Advanced Institute of Science and Technology*)

Two-part Segmentation of Text Documents (Page 793)

Deepak P (*IBM Research - India*)
 Karthik Viswesvariah (*IBM Research - India*)
 Nimalie Wiratunga (*Robert Gordon University*)
 Sadiq Sani (*Robert Gordon University*)

On the Design of LDA Models for Aspect-based Opinion Mining (Page 803)

Samaneh Moghaddam (*Simon Fraser University*)
 Martin Ester (*Simon Fraser University*)

Session 18 - IR Track: Formal Retrieval Models and Learning to Rank

Predicting Query Performance for Fusion-Based Retrieval (Page 813)

Gad Markovits (*Technion - Israel Institute of Technology*)
 Anna Shtok (*Technion - Israel Institute of Technology*)
 Oren Kurland (*Technion - Israel Institute of Technology*)
 David Carmel (*IBM Research lab*)

Back to the Roots: A Probabilistic Framework for Query-Performance Prediction (Page 823)

Oren Kurland (*Technion - Israel Institute of Technology*)
 Anna Shtok (*Technion - Israel Institute of Technology*)
 Shay Hummel (*Technion - Israel Institute of Technology*)
 Fiana Raiber (*Technion - Israel Institute of Technology*)
 David Carmel (*IBM Research lab*)
 Ofri Rom (*Technion - Israel Institute of Technology*)

Learning to Rank for Robust Question Answering (Page 833)

Arvind Agarwal (*University of Maryland*)
 Hema Raghavan (*IBM T.J. Watson Research Center*)
 Karthik Subbian (*University of Minnesota*)
 Prem Melville (*IBM T.J. Watson Research Center*)
 Richard D. Lawrence (*IBM T.J. Watson Research Center*)
 David C. Gondek (*IBM T.J. Watson Research Center*)
 James Fan (*IBM T.J. Watson Research Center*)

Learning to Rank by Aggregating Expert Preferences (Page 843)

Maksims N. Volkovs (*University of Toronto*)
 Hugo Larochelle (*Université de Sherbrooke*)
 Richard S. Zemel (*University of Toronto*)

Learning to Rank Duplicate Bug Reports (Page 852)

Jian Zhou (*Tsinghua University*)
 Hongyu Zhang (*Tsinghua University*)

Session 19 - DB Track: Probabilistic and Uncertain Data

A Model-based Approach for RFID Data Stream Cleansing (Page 862)

Zhou Zhao (*The Hong Kong University of Science and Technology*)

Wilfred Ng (*The Hong Kong University of Science and Technology*)

What is the IQ of Your Data Transformation System? (Page 872)

Giansalvatore Mecca (*Università della Basilicata*)

Paolo Papotti (*Qatar Computing Research Institute*)

Salvatore Raunich (*University of Leipzig*)

Donatello Santoro (*Università della Basilicata & Università Roma Tre*)

(Return to Top)

On the Foundations of Probabilistic Information Integration (Page 882)

Fereidoon Sadri (*University of North Carolina at Greensboro*)

GPU Acceleration of Probabilistic Frequent Itemset Mining from Uncertain Databases (Page 892)

Yusuke Kozawa (*University of Tsukuba*)

Toshiyuki Amagasa (*University of Tsukuba & Japan Aerospace Exploration Agency*)

Hiroyuki Kitagawa (*University of Tsukuba*)

Completeness of Queries over SQL Databases (Page 902)

Werner Nutt (*Free University of Bozen-Bolzano*)

Simon Razniewski (*Free University of Bozen-Bolzano*)

Session 20 - DB Track: Top-k and Nearest Neighbor Queries

Being Picky-Processing Top-K Queries with Set-Defined Selections (Page 912)

Aleksandar Stupar (*Saarland University*)

Sebastian Michel (*Saarland University*)

Finding Top k Most Influential Spatial Facilities over Uncertain Objects (Page 922)

Liming Zhan (*The University of New South Wales*)

Ying Zhang (*The University of New South Wales*)

Wenjie Zhang (*The University of New South Wales*)

Xuemin Lin (*The University of New South Wales*)

(Return to Top)

Efficient Safe-Region Construction for Moving Top-K Spatial Keyword Queries (Page 932)

Weihuang Huang (*Tsinghua University*)

Guoliang Li (*Tsinghua University*)

Kian-Lee Tan (*National University of Singapore*)

Jianhua Feng (*Tsinghua University*)

Monochromatic and Bichromatic Reverse Nearest Neighbor Queries on Land Surfaces (Page 942)

Da Yan (*The Hong Kong University of Science and Technology*)

Zhou Zhao (*The Hong Kong University of Science and Technology*)

Wilfred Ng (*The Hong Kong University of Science and Technology*)

Pay-as-you-go Maintenance of Precomputed Nearest Neighbors in Large Graphs (Page 952)

Tom Crecelius (*Max Planck Institute for Informatics*)

Ralf Schenkel (*Saarland University*)

Session 21 - KM Track: Spatial and Temporal Methods

Spatial Influence vs. Community Influence: Modeling the Global Spread of Social Media (Page 962)

Krishna Y. Kamath (*Texas A&M University*)

James Caverlee (*Texas A&M University*)

Zhiyuan Cheng (*Texas A&M University*)

Daniel Z. Sui (*Ohio State University*)

TUT: A Statistical Model for Detecting Trends, Topics and User Interests in Social Media (Page 972)

Xuning Tang (*Drexel University*)

Christopher C. Yang (*Drexel University*)

(Return to Top)

Predicting Aggregate Social Activities using Continuous-Time Stochastic Process (Page 982)

Shu Huang (*The Pennsylvania State University*)

Min Chen (*Johns Hopkins University*)

Bo Luo (*The University of Kansas*)

Dongwon Lee (*The Pennsylvania State University*)

Acquiring Temporal Constraints between Relations (Page 992)

Partha Pratim Talukdar (*Carnegie Mellon University*)

Derry Wijaya (*Carnegie Mellon University*)

Tom Mitchell (*Carnegie Mellon University*)

Session 22 - IR Track: Web Search

Towards Optimum Query Segmentation: In Doubt Without (Page 1015)

Matthias Hagen (*Bauhaus-Universität Weimar*)

Martin Potthast (*Bauhaus-Universität Weimar*)

Anna Beyer (*Bauhaus-Universität Weimar*)

Benno Stein (*Bauhaus-Universität Weimar*)

Leaving So Soon? Understanding and Predicting Web Search Abandonment Rationales (Page 1025)

Leaving So Soon? Understanding and Predicting Web Search Abandonment Rationales (Page 1029)

Abdigani Diriye (*University College London*)

Ryen W. White (*Microsoft Corporation*)

Georg Buscher (*Microsoft Corporation*)

Susan T. Dumais (*Microsoft Corporation*)

(Return to Top)

Click Patterns: An Empirical Representation of Complex Query Intents (Page 1035)

Huizhong Duan (*University of Illinois at Urbana-Champaign*)

Emre Kiciman (*Microsoft Research*)

ChengXiang Zhai (*University of Illinois at Urbana-Champaign*)

Domain Dependent Query Reformulation for Web Search (Page 1045)

Van Dang (*University of Massachusetts, Amherst*)

Giridhar Kumaran (*Microsoft Corporation*)

Adam Troy (*Microsoft Corporation*)

Session 23 - DB Track: Web Data Management

An Automatic Blocking Mechanism for Large-Scale De-Duplication Tasks (Page 1055)

Anish Das Sarma (*Google Research*)

Ankur Jain (*Yahoo! Research*)

Ashwin Machanavajjhala (*Duke University*)

Philip Bohannon (*Facebook*)

Processing Continuous Text Queries Featuring Non-Homogeneous Scoring Functions (Page 1065)

Nelly Vouzoukidou (*Pierre et Marie Curie University*)

Bernd Amann (*Pierre et Marie Curie University*)

Vassilis Christophides (*ICS/FORTH & University of Crete*)

(Return to Top)

Comprehension-Based Result Snippets (Page 1075)

Abhijith Kashyap (*University of California, Riverside*)

Vagelis Hristidis (*University of California, Riverside*)

An Effective Rule Miner for Instance Matching in a Web of Data (Page 1085)

Xing Niu (*Shanghai Jiao Tong University*)

Shu Rong (*Shanghai Jiao Tong University*)

Haofen Wang (*Shanghai Jiao Tong University*)

Yong Yu (*Shanghai Jiao Tong University*)

Session 24 - KM Track: Information Extraction

Non-Stationary Bayesian Networks Based on Perfect Simulation (Page 1095)

Yi Jia (*University of Kansas*)

Wenrong Zeng (*University of Kansas*)

Jun Huan (*University of Kansas*)

Active Learning for Relation Type Extension with Local and Global Data Views (Page 1105)

Ang Sun (*New York University*)

Ralph Grishman (*New York University*)

Segmenting Web-Domains and Hashtags Using Length Specific Models (Page 1113)

Sriram Srinivasan (*Yahoo! SDC*)

Sourangshu Bhattacharya (*Yahoo! Labs*)

Rudras Chakraborty (*Indian Statistical Institute*)

(Return to Top)

Crosslingual Distant Supervision for Extracting Relations of Different Complexity (Page 1123)

Andre Blessing (*Universität Stuttgart*)

Hinrich Schütze (*Universität Stuttgart*)

Labeling by Landscaping: Classifying Tokens in Context by Pruning and Decorating Trees (Page 1133)

Siddharth Patwardhan (*IBM T.J. Watson Research Center*)

Branimir Boguraev (*IBM T.J. Watson Research Center*)

Apoorv Agarwal (*Columbia University*)

Alessandro Moschitti (*University of Trento*)

Jennifer Chu-Carroll (*IBM T.J. Watson Research Center*)

Session 25 - IR Track: Topic Modeling and Content and Sentiment Analysis

G-WSTD: A Framework for Geographic Web Search Topic Discovery (Page 1143)

Di Jiang (*Hong Kong University of Science and Technology*)

Jan Vosecky (*Hong Kong University of Science and Technology*)

Kenneth Wai-Ting Leung (*Hong Kong University of Science and Technology*)

Wilfred Ng (*Hong Kong University of Science and Technology*)

Supporting Factual Statements with Evidence from the Web (Page 1153)

Chee Wee Leong (*University of North Texas*)

Silviu Cucerzan (*Microsoft Research*)

(Return to Top)

Role-Explicit Query Identification and Intent Role Annotation (Page 1163)

Haitao Yu (*University of Tokushima*)

Fuji Ren (*University of Tokushima*)

Joint Topic Modeling for Event Summarization Across News and Social Media Streams (Page 1173)

Wei Gao (*Qatar Computing Research Institute*)
 Peng Li (*Shanghai Jiaotong University*)
 Kareem Darwish (*Qatar Computing Research Institute*)

Session 26 - DB Track: Query Processing, Optimization and Performance

CGStream: Continuous Correlated Graph Query for Data Streams (Page 1183)

Shirui Pan (*University of Technology, Sydney*)
 Xingquan Zhu (*University of Technology, Sydney*)

Efficient Influence-Based Processing of Market Research Queries (Page 1193)

Anastasios Arvanitis (*National Technical University of Athens*)
 Antonios Deligiannakis (*Technical University of Crete*)
 Yannis Vassiliou (*National Technical University of Athens*)

Deco: Declarative Crowdsourcing (Page 1203)

Aditya Parameswaran (*Stanford University*)
 Hyunjung Park (*Stanford University*)
 Hector Garcia-Molina (*Stanford University*)
 Neoklis Polyzotis (*University of California, Santa Cruz*)
 Jennifer Widom (*Stanford University*)

Predicting the Effectiveness of Keyword Queries on Databases (Page 1213)

Shiwen Cheng (*University of California, Riverside*)
 Arash Termehchy (*University of Illinois at Urbana-Champaign*)
 Vagelis Hristidis (*University of California, Riverside*)

You Can Stop Early with COLA: Online Processing of Aggregate Queries in the Cloud (Page 1223)

Yingjie Shi (*Renmin University of China*)
 Xiaofeng Meng (*Renmin University of China*)
 Fusheng Wang (*Emory University*)
 Yantao Gan (*Renmin University of China*)

Session 27 - KM Track: Classification and Semantic Methods

A Novel Local Patch Framework for Fixing Supervised Learning Models (Page 1233)

Yilei Wang (*Peking University*)
 Bingzheng Wei (*Microsoft Research Asia*)
 Jun Yan (*Microsoft Research Asia*)
 Yang Hu (*Shanghai Jiao Tong University*)
 Zhi-Hong Deng (*Peking University*)
 Zheng Chen (*Microsoft Research Asia*)

Automated Feature Weighting in Naive Bayes for High-Dimensional Data Classification (Page 1243)

Lifei Chen (*University of Sherbrooke*)
 Shengrui Wang (*University of Sherbrooke*)

Learning to Discover Complex Mappings from Web Forms to Ontologies (Page 1253)

Yuan An (*Drexel University*)
 Xiaohua Hu (*Drexel University*)
 Il-Yeol Song (*Drexel University*)

Modeling Semantic Relations Between Visual Attributes and Object Categories via Dirichlet Forest Prior (Page 1263)

Xin Chen (*Drexel University*)
 Xiaohua Hu (*Drexel University*)
 Zhongna Zhou (*University of Missouri - Columbia*)
 Yuan An (*Drexel University*)
 Tingting He (*Central China Normal University*)
 E. K. Park (*California State University - Chico*)

CoNet: Feature Generation for Multi-View Semi-Supervised Learning with Partially Observed Views (Page 1273)

Brian Quanz (*University of Kansas*)
 Jun Huan (*University of Kansas*)

Session 28 - IR Track: Multimedia and User Feedback

Generating Facets for Phone-Based Navigation of Structured Data (Page 1283)

Krishna Kumamuru (*IBM*)
 Ajith Jujjuru (*IBM*)
 Mayuri Duggirala (*IBM*)

The Effect of Aggregated Search Coherence on Search Behavior (Page 1293)

Jaime Arguello (*University of North Carolina at Chapel Hill*)
 Robert Capra (*University of North Carolina at Chapel Hill*)

Improving Bag-of-visual-Words Model with Spatial-Temporal Correlation for Video Retrieval (Page 1303)

Lei Wang (*Robert Gordon University*)
 Dawei Song (*Tianjin University & The Open University*)
 Eyad Elyan (*Robert Gordon University*)

Exploring and Predicting Search Task Difficulty (Page 1313)

Jingjing Liu (*University of South Carolina*)
 Chang Liu (*Rutgers University*)

Michael Cole (*Rutgers University*)
 Nicholas J. Belkin (*Rutgers University*)
 Xiangmin Zhang (*Wayne State University*)

Iterative Relevance Feedback with Adaptive Exploration/Exploitation Trade-off (Page 1323)
 Nicolae Suditu (*Idiap Research Institute & École Polytechnique Fédérale de Lausanne*)
 François Fleuret (*Idiap Research Institute & École Polytechnique Fédérale de Lausanne*)

Session 29 - DB Track: Emerging and Advanced Topics

A Practical Concurrent Index for Solid-State Drives (Page 1332)

Risi Thonangi (*Duke University*)
 Shivnath Babu (*Duke University*)
 Jun Yang (*Duke University*)

Robust Distributed Indexing for Locality-Skewed Workloads (Page 1342)

Mu-Woong Lee (*Pohang University of Science and Technology*)
 Seung-won Hwang (*Pohang University of Science and Technology*)

Efficient Provenance Storage for Relational Queries (Page 1352)

Zhifeng Bao (*National University of Singapore*)
 Henning Köhler (*University of Queensland*)
 Liwei Wang (*Wuhan University*)
 Xiaofang Zhou (*University of Queensland*)
 Shazia Sadiq (*University of Queensland*)

Generically Extending Anonymization Algorithms to Deal with Successive Queries (Page 1362)

Manuel Barbosa (*HASLab-INESC TEC & Universidade do Minho*)
 Alexandre Pinto (*HASLab-INESC TEC & Instituto Superior da Maia*)
 Bruno Gomes (*HASLab-INESC TEC & Universidade do Minho*)

Authentication of Moving Range Queries (Page 1372)

Duncan Yung (*Hong Kong Polytechnic University*)
 Eric Lo (*Hong Kong Polytechnic University*)
 Man Lung Yiu (*Hong Kong Polytechnic University*)

Session 30 - KM Track: Novel Applications

Model the Complex Dependence Structures of Financial Variables by Using Canonical Vine (Page 1382)

Wei Wei (*University of Technology, Sydney*)
 Xuhui Fan (*University of Technology, Sydney*)
 Jinyan Li (*University of Technology, Sydney*)
 Longbing Cao (*University of Technology, Sydney*)

A Unified Learning Framework for Auto Face Annotation by Mining Web Facial Images (Page 1392)

Dayong Wang (*Nanyang Technological University*)
 Steven C. H. Hoi (*Nanyang Technological University*)
 Ying He (*Nanyang Technological University*)

Efficient Jaccard-Based Diversity Analysis of Large Document Collections (Page 1402)

Fan Deng (*University of Hannover*)
 Stefan Siersdorfer (*University of Hannover*)
 Sergej Zerr (*University of Hannover*)

Knowing Where and How Criminal Organizations Operate Using Web Content (Page 1412)

Michele Coscia (*KddLab ISTI CNR*)
 Viridiana Rios (*Harvard University*)

Session 31 - IR Track: Social Networks

Social Recommendation Across Multiple Relational Domains (Page 1422)

Meng Jiang (*Tsinghua University*)
 Peng Cui (*Tsinghua University*)
 Fei Wang (*IBM T.J. Watson Research Center*)
 Qiang Yang (*Hong Kong University of Science and Technology*)
 Wenwu Zhu (*Tsinghua University*)
 Shiqiang Yang (*Tsinghua University*)

Mining Competitive Relationships by Learning Across Heterogeneous Networks (Page 1432)

Yang Yang (*Tsinghua University*)
 Jie Tang (*Tsinghua University*)
 Jacklyne Keomany (*Tsinghua University*)
 Yanting Zhao (*Tsinghua University*)

Juanzi Li (*Tsinghua University*)
 Ying Ding (*Indiana University*)
 Tian Li (*Tsinghua University*)
 Liangwei Wang (*Huawei Noah's Ark Lab*)

Evaluating Geo-Social Influence in Location-Based Social Networks (Page 1442)

Chao Zhang (*Zhejiang University*)
 Lidan Shou (*Zhejiang University*)

(Return to Top)

(Return to Top)

(Return to Top)

Ke Chen (*Zhejiang University*)
 Gang Chen (*Zhejiang University*)
 Yijun Bei (*Zhejiang University*)

The Walls Have Ears: Optimize Sharing for Visibility and Privacy in Online Social Networks (Page 1452)

Thang N. Dinh (*University of Florida*)
 Yilin Shen (*University of Florida*)
 My T. Thai (*University of Florida*)

Knowledge Management Short Paper Session

Influence and Similarity on Heterogeneous Networks (Page 1462)

Guan Wang (*University of Illinois at Chicago*)
 Qingbo Hu (*University of Illinois at Chicago*)
 Philip S. Yu (*University of Illinois at Chicago & King Abdulaziz University*)

GRAFT: An Approximate Graphlet Counting Algorithm for Large Graph Analysis (Page 1467)

Mahmudur Rahman (*Indiana University - Purdue University, Indianapolis*)
 Mansurul Bhuiyan (*Indiana University - Purdue University, Indianapolis*)
 Mohammad Al Hasan (*Indiana University - Purdue University, Indianapolis*)

Hierarchical Co-Clustering Based on Entropy Splitting (Page 1472)

Wei Cheng (*University of North Carolina at Chapel Hill*)
 Xiang Zhang (*Case Western Reserve University*)
 Feng Pan (*Microsoft*)
 Wei Wang (*University of California, Los Angeles*)

Mining Long-Lasting Exploratory User Interests from Search History (Page 1477)

Bin Tan (*University of Illinois at Urbana-Champaign*)
 Yuanhua Lv (*University of Illinois at Urbana-Champaign*)
 ChengXiang Zhai (*University of Illinois at Urbana-Champaign*)

Feature Selection Based on Term Frequency and T-Test for Text Categorization (Page 1482)

Deqing Wang (*Beihang University*)
 Hui Zhang (*Beihang University*)
 Rui Liu (*Beihang University*)
 Weifeng Lv (*Beihang University*)

Adapting Vector Space Model to Ranking-Based Collaborative Filtering (Page 1487)

Shuaiqiang Wang (*Shandong University of Finance and Economics*)
 Jiankai Sun (*Shandong University*)
 Byron J. Gao (*Texas State University-San Marcos*)
 Jun Ma (*Shandong University*)

Joint Relevance and Answer Quality Learning for Question Routing in Community QA (Page 1492)

Guangyou Zhou (*Chinese Academy of Sciences*)
 Kang Liu (*Chinese Academy of Sciences*)
 Jun Zhao (*Chinese Academy of Sciences*)

Fast Approximation of Steiner Trees in Large Graphs (Page 1497)

Andrey Gubichev (*Technische Universität München*)
 Thomas Neumann (*Technische Universität München*)

Automatically Embedding Newsworthy Links to Articles (Page 1502)

Hakan Ceylan (*Netflix*)
 Ioannis Arapakis (*Yahoo! Labs*)
 Pinar Donmez (*Salesforce.com*)
 Mounia Lalmas (*Yahoo! Labs*)

Learning Spectral Embedding via Iterative Eigenvalue Thresholding (Page 1507)

Fanhua Shang (*Xidian University*)
 L. C. Jiao (*Xidian University*)
 Yuanyuan Liu (*Xidian University*)
 Fei Wang (*IBM T.J. Watson Research Center*)

Measuring Robustness of Complex Networks Under MVC Attack (Page 1512)

Rong-Hua Li (*The Chinese University of Hong Kong*)
 Jeffrey Xu Yu (*The Chinese University of Hong Kong*)
 Xin Huang (*The Chinese University of Hong Kong*)
 Hong Cheng (*The Chinese University of Hong Kong*)
 Zechao Shang (*The Chinese University of Hong Kong*)

A Simple Approach to the Design of Site-Level Extractors Using Domain-Centric Principles (Page 1517)

Chong Long (*Yahoo! Labs*)
 Xiubo Geng (*Yahoo! Labs*)
 Chang Xu (*Chinese Academy of Sciences*)
 Sathya Keerthi (*Microsoft*)

Extraction of Topic Evolutions from References in Scientific Articles and Its GPU Acceleration (Page 1522)

Tomonari Masada (*Nagasaki University*)
 Atsuhiko Takasu (*National Institute of Informatics*)

Graph-Based Workflow Recommendation: On Improving Business Process Modeling (Page 1527)

(Return to Top)

(Return to Top)

[Graph-Based Network Recommendation: On Improving Business Process Modeling](#) (Page 1521)

Bin Cao (*Zhejiang University*)
 Jianwei Yin (*Zhejiang University*)
 Shuiguang Deng (*Zhejiang University*)
 Dongjing Wang (*Zhejiang University*)
 Zhaohui Wu (*Zhejiang University*)

Reconciling Ontologies with the Web of Data (Page 1532)

Ziawasch Abedjan (*Hasso-Plattner-Institut*)
 Johannes Lorey (*Hasso-Plattner-Institut*)
 Felix Naumann (*Hasso-Plattner-Institut*)

Efficient Extraction of Ontologies from Domain Specific Text Corpora (Page 1537)

Tianyu Li (*University of British Columbia*)
 Pirooz Chubak (*University of British Columbia*)
 Laks V.S. Lakshmanan (*University of British Columbia*)
 Rachel Pottinger (*University of British Columbia*)

Effective and Efficient? Bilingual Sentiment Lexicon Extraction Using Collocation Alignment (Page 1542)

Zheng Lin (*Chinese Academy of Sciences*)
 Songbo Tan (*Chinese Academy of Sciences*)
 Xueqi Cheng (*Chinese Academy of Sciences*)
 Xueke Xu (*Chinese Academy of Sciences*)
 Weisong Shi (*Wayne State University*)

[\(Return to Top\)](#)

Exploiting Latent Relevance for Relational Learning of Ubiquitous Things (Page 1547)

Lina Yao (*The University of Adelaide*)
 Quan Z. Sheng (*The University of Adelaide*)

Discovering Personally Semantic Places from GPS Trajectories (Page 1552)

Mingqi Lv (*Zhejiang University*)
 Ling Chen (*Zhejiang University*)
 Gencai Chen (*Zhejiang University*)

Mining Coherent Anomaly Collections on Web Data (Page 1557)

Hanbo Dai (*Singapore Management University*)
 Feida Zhu (*Singapore Management University*)
 Ee-Peng Lim (*Singapore Management University*)
 HweeHwa Pang (*Singapore Management University*)

Mining Topic-Level Opinion Influence in Microblog (Page 1562)

Daifeng Li (*Tsinghua University*)
 Xin Shuai (*Indiana University Bloomington*)
 Guozheng Sun (*Tencent Company*)
 Jie Tang (*Tsinghua University*)
 Ying Ding (*Indiana University Bloomington*)
 Zhipeng Luo (*Beijing University of Aeronautics and Astronautics*)

Meta Path-Based Collective Classification in Heterogeneous Information Networks (Page 1567)

Xiangnan Kong (*University of Illinois at Chicago*)
 Philip S. Yu (*University of Illinois at Chicago & King Abdulaziz University*)
 Ying Ding (*Indiana University Bloomington*)
 David J. Wild (*Indiana University Bloomington*)

Discretionary Social Network Data Revelation with a User-Centric Utility Guarantee (Page 1572)

Yi Song (*National University of Singapore*)
 Giorgos Cheliotis (*National University of Singapore*)
 Panagiotis Karras (*Rutgers University*)
 Mingqiang Xue (*Institute for Infocomm Research*)
 Sadegh Nobari (*National University of Singapore*)
 Stéphane Bressan (*National University of Singapore*)

[\(Return to Top\)](#)

Empirical Validation of the Buckley-Osthus Model for the Web Host Graph: Degree and Edge Distributions (Page 1577)

Maxim Zhukovskiy (*Yandex*)
 Dmitry Vinogradov (*Yandex*)
 Yuri Pritykin (*Yandex*)
 Liudmila Ostroumova (*Yandex*)
 Evgeniy Grechnikov (*Yandex*)
 Gleb Gusev (*Yandex*)
 Pavel Serdyukov (*Yandex*)
 Andrei Raigorodskii (*Yandex*)

gSCorr: Modeling Geo-Social Correlations for New Check-ins on Location-Based Social Networks (Page 1582)

Huiji Gao (*Arizona State University*)
 Jiliang Tang (*Arizona State University*)
 Huan Liu (*Arizona State University*)

Swimming Against the Streamz: Search and Analytics Over the Enterprise Activity Stream (Page 1587)

Ido Guy (*IBM Research*)
 Tal Steier (*IBM Research*)
 Maya Barnea (*IBM Research*)
 Inbal Ronen (*IBM Research*)
 Tal Daniel (*IBM Research*)

What Is Happening Right Now ... That Interests Me?: Online Topic Discovery and Recommendation in Twitter (Page 1592)

Ernesto Diaz-Aviles (*University of Hannover*)
 Lucas Drumond (*University of Hildesheim*)
 Zeno Gantner (*University of Hildesheim*)
 Lars Schmidt-Thieme (*University of Hildesheim*)
 Wolfgang Nejdl (*University of Hannover*)

Frequent grams Based Embedding for Privacy Preserving Record Linkage (Page 1597)

Luca Bonomi (*Emory University*)
 Li Xiong (*Emory University*)
 Rui Chen (*Concordia University*)
 Benjamin C. M. Fung (*Concordia University*)

If You Are Happy and You Know It... Tweet (Page 1602)

Amir Asiaee T. (*University of Minnesota*)
 Mariano Tepper (*Duke University*)
 Arindam Banerjee (*University of Minnesota*)
 Guillermo Sapiro (*Duke University*)

[\(Return to Top\)](#)**PRemISE: Personalized News Recommendation via Implicit Social Experts** (Page 1607)

Chen Lin (*Xiamen University*)
 Runquan Xie (*Xiamen University*)
 Lei Li (*Florida International University*)
 Zhenhua Huang (*Tongji University*)
 Tao Li (*Florida International University*)

Hierarchical Topic Integration Through Semi-Supervised Hierarchical Topic Modeling (Page 1612)

Xian-Ling Mao (*Peking University*)
 Jing He (*Université de Montréal*)
 Hongfei Yan (*Peking University*)
 Xiaoming Li (*Peking University*)

Exploiting Enriched Contextual Information for Mobile App Classification (Page 1617)

Hengshu Zhu (*University of Science and Technology of China*)
 Huanhuan Cao (*Nokia Research Center*)
 Enhong Chen (*University of Science and Technology of China*)
 Hui Xiong (*Rutgers University*)
 Jilei Tian (*Nokia Research Center*)

Incorporating Word Correlation into Tag-Topic Model for Semantic Knowledge Acquisition (Page 1622)

Fang Li (*Central China Normal University*)
 Tingting He (*Central China Normal University*)
 Xinhui Tu (*Central China Normal University*)
 Xiaohua Hu (*Central China Normal University / Drexel University*)

PriSM: Discovering and Prioritizing Severe Technical Issues from Product Discussion Forums (Page 1627)

Rashmi Gangadharaiah (*IBM Research India*)
 Rose Catherine (*IBM Research India*)

Preprocessing of Informal Mathematical Discourse in Context of Controlled Natural Language (Page 1632)

Raúl E. Gutiérrez de Piñeres Reyes (*Universidad del Valle*)
 Juan Francisco Díaz Frias (*Universidad del Valle*)

PathRank: A Novel Node Ranking Measure on a Heterogeneous Graph for Recommender Systems (Page 1637)

Sangkeun Lee (*Seoul National University*)
 Sungchan Park (*Seoul National University*)
 Minsuk Kahng (*Seoul National University*)
 Sang-goo Lee (*Seoul National University*)

[\(Return to Top\)](#)**Unsupervised Discovery of Opposing Opinion Networks from Forum Discussions** (Page 1642)

Yue Lu (*Twitter Inc.*)
 Hongning Wang (*University of Illinois at Urbana-Champaign*)
 ChengXiang Zhai (*University of Illinois at Urbana-Champaign*)
 Dan Roth (*University of Illinois at Urbana-Champaign*)

Exploring the Existing Category Hierarchy to Automatically Label the Newly-Arising Topics in cQA (Page 1647)

Guangyou Zhou (*Chinese Academy of Sciences*)
 Li Cai (*Chinese Academy of Sciences*)
 Kang Liu (*Chinese Academy of Sciences*)
 Jun Zhao (*Chinese Academy of Sciences*)

Query-Focused Multi-Document Summarization Based on Query-Sensitive Feature Space (Page 1652)

Wenpeng Yin (*Peking University Shenzhen Graduate School*)
 Yulong Pei (*Peking University Shenzhen Graduate School*)
 Fan Zhang (*Peking University Shenzhen Graduate School*)
 Lian'en Huang (*Peking University Shenzhen Graduate School*)

Time-Aware Topic Recommendation Based on Micro-Blogs (Page 1657)

Huizhi Liang (*The Australian National University*)
 Yue Xu (*Queensland University of Technology*)
 Dian Tjondronegoro (*Queensland University of Technology*)
 Peter Christen (*The Australian National University*)

Peter Christen (*The Australian National University*)

Topic-Sensitive Probabilistic Model for Expert Finding in Question Answer Communities (Page 1662)

Guangyou Zhou (*Chinese Academy of Sciences*)

Siwei Lai (*Chinese Academy of Sciences*)

Kang Liu (*Chinese Academy of Sciences*)

Jun Zhao (*Chinese Academy of Sciences*)

iSampling: Framework for Developing Sampling Methods Considering User's Interest (Page 1667)

Jinoh Oh (*Pohang University of Science and Technology*)

Hwanjo Yu (*Pohang University of Science and Technology*)

WiSeNet: Building a Wikipedia-Based Semantic Network with Ontologized Relations (Page 1672)

Andrea Moro (*Università di Roma*)

Roberto Navigli (*Università di Roma*)

Shaping Communities out of Triangles (Page 1677)

Arnau Prat-Pérez (*Universitat Politècnica de Catalunya*)

David Dominguez-Sal (*Universitat Politècnica de Catalunya*)

Josep M. Brunat (*Universitat Politècnica de Catalunya*)

Josep-Lluís Larriba-Pey (*Universitat Politècnica de Catalunya*)

The Early-Adopter Graph and Its Application to Web-Page Recommendation (Page 1682)

Ida Mele (*University of Rome*)

Francesco Bonchi (*Yahoo! Research*)

Aristides Gionis (*Yahoo! Research*)

Relational Co-Clustering via Manifold Ensemble Learning (Page 1687)

Ping Li (*Zhejiang University*)

Jiajun Bu (*Zhejiang University*)

Chun Chen (*Zhejiang University*)

Zhanying He (*Zhejiang University*)

SemaFor: Semantic Document Indexing Using Semantic Forests (Page 1692)

George Tsatsaronis (*Technische Universität Dresden*)

Iraklis Varlamis (*Harokopio University of Athens*)

Kjetil Nørvgård (*Norwegian University of Science and Technology*)

Measuring Website Similarity Using an Entity-Aware Click Graph (Page 1697)

Pablo N. Mendes (*Freie Universität Berlin*)

Peter Mika (*Yahoo! Research Barcelona*)

Hugo Zaragoza (*Yahoo! Research Barcelona*)

Roi Blanco (*Yahoo! Research Barcelona*)

Community-Based Classification of Noun Phrases in Twitter (Page 1702)

Freddy Chong Tat Chua (*Singapore Management University*)

William W. Cohen (*Carnegie Mellon University*)

Justin Betteridge (*Carnegie Mellon University*)

Ee-Peng Lim (*Singapore Management University*)

Real-Time Bid Optimization for Group-Buying Ads (Page 1707)

Raju Balakrishnan (*Groupon Inc.*)

Rushi P. Bhatt (*Amazon.com*)

Degree Relations of Triangles in Real-World Networks and Graph Models (Page 1712)

Nurcan Durak (*Sandia National Laboratories*)

Ali Pinar (*Sandia National Laboratories*)

Tamara G. Kolda (*Sandia National Laboratories*)

C. Seshadhri (*Sandia National Laboratories*)

A Probabilistic Approach to Mining Geospatial Knowledge from Social Annotations (Page 1717)

Suradej Intagorn (*University of Southern California*)

Kristina Lerman (*University of Southern California*)

Providing Grades and Feedback for Student Summaries by Ontology-Based Information Extraction (Page 1722)

Fernando Gutierrez (*University of Oregon*)

Dejing Dou (*University of Oregon*)

Stephen Fickas (*University of Oregon*)

Gina Griffiths (*University of Oregon*)

Joint Bilingual Name Tagging for Parallel Corpora (Page 1727)

Qi Li (*City University of New York*)

Haibo Li (*Queens College & City University of New York*)

Heng Ji (*Queens College & City University of New York*)

Wen Wang (*SRI International*)

Jing Zheng (*SRI International*)

Fei Huang (*IBM T.J. Watson Research Center*)

Using Program Synthesis for Social Recommendations (Page 1732)

Alvin Cheung (*Massachusetts Institute of Technology*)

Armando Solar-Lezama (*Massachusetts Institute of Technology*)

Samuel Madden (*Massachusetts Institute of Technology*)

(Return to Top)

(Return to Top)

Web-Scale Multi-Task Feature Selection for Behavioral Targeting (Page 1737)

Amr Ahmed (*Google Research*)
 Mohamed Aly (*Yahoo! Research*)
 Abhimanyu Das (*Yahoo! Research*)
 Alexander J. Smola (*Google Research*)
 Tasos Anastasakos (*Yahoo! Research*)

Balanced Coverage of Aspects for Text Summarization (Page 1742)

Takuya Makino (*Fujitsu Laboratories Ltd.*)
 Hiroya Takamura (*Tokyo Institute of Technology*)
 Manabu Okumura (*Tokyo Institute of Technology*)

Dynamic Effects of Ad Impressions on Commercial Actions in Display Advertising (Page 1747)

Joel Barajas (*University of California, Santa Cruz*)
 Ram Akella (*University of California, Santa Cruz*)
 Marius Holtan (*AOL Research*)
 Jaimie Kwon (*AOL Research*)
 Aaron Flores (*AOL Research*)
 Victor Andrei (*AOL Research*)

[\(Return to Top\)](#)**A Hybrid Approach for Efficient Provenance Storage** (Page 1752)

Yulai Xie (*Huazhong University of Science and Technology*)
 Dan Feng (*Huazhong University of Science and Technology*)
 Zhipeng Tan (*Huazhong University of Science and Technology*)
 Lei Chen (*Huazhong University of Science and Technology*)
 Kiran-Kumar Muniswamy-Reddy (*Harvard University*)
 Yan Li (*University of California, Santa Cruz*)
 Darrell D. E. Long (*University of California, Santa Cruz*)

Information Retrieval Short Paper Session**Content-Based Relevance Estimation on the Web using Inter-Document Similarities** (Page 1769)

Fiana Raiber (*Technion*)
 Oren Kurland (*Technion*)
 Moshe Tennenholtz (*Microsoft Research and Technion*)

Trust Prediction via Aggregating Heterogeneous Social Networks (Page 1774)

Jin Huang (*University of Texas at Arlington*)
 Feiping Nie (*University of Texas at Arlington*)
 Heng Huang (*University of Texas at Arlington*)
 Yi-Cheng Tu (*University of South Florida*)

[\(Return to Top\)](#)**Estimating Interleaved Comparison Outcomes from Historical Click Data** (Page 1779)

Katja Hofmann (*University of Amsterdam*)
 Shimon Whiteson (*University of Amsterdam*)
 Maarten de Rijke (*University of Amsterdam*)

Automatic Image Annotation Using Tag-related Random Search over Visual Neighbors (Page 1784)

Zijia Lin (*Tsinghua University*)
 Guiguang Ding (*Tsinghua University*)
 Mingqing Hu (*Chinese Academy of Sciences*)
 Jianmin Wang (*Tsinghua University*)
 Jianguang Sun (*Tsinghua University*)

Diversionsary Comments Under Political Blog Posts (Page 1789)

Jing Wang (*University of Illinois at Chicago*)
 Clement T. Yu (*University of Illinois at Chicago*)
 Philip S. Yu (*University of Illinois at Chicago*)
 Bing Liu (*University of Illinois at Chicago*)
 Weiyi Meng (*SUNY at Binghamton*)

Discover Breaking Events with Popular Hashtags in Twitter (Page 1794)

Anqi Cui (*Tsinghua University*)
 Min Zhang (*Tsinghua University*)
 Yiqun Liu (*Tsinghua University*)
 Shaoping Ma (*Tsinghua University*)
 Kuo Zhang (*Tsinghua University*)

[\(Return to Top\)](#)**Query Likelihood with Negative Query Generation** (Page 1799)

Yuanhua Lv (*University of Illinois at Urbana-Champaign*)
 ChengXiang Zhai (*University of Illinois at Urbana-Champaign*)

On the Connections between Explicit Semantic Analysis and Latent Semantic Analysis (Page 1804)

Chao Liu (*Tencent Inc*)
 Yi-Min Wang (*Microsoft Research*)

Variance Maximization via Noise Injection for Active Sampling in Learning to Rank (Page 1809)

Wenbin Cai (*Shanghai Jiao Tong University*)
 Ya Zhang (*Shanghai Jiao Tong University*)

More Than Relevance: High Utility Query Recommendation by Mining Users' Search Behaviors (Page 1814)

Xiaofei Zhu (*Chinese Academy of Sciences*)

Jiafeng Guo (*Chinese Academy of Sciences*)
 Xueqi Cheng (*Chinese Academy of Sciences*)
 Yanyan Lan (*Chinese Academy of Sciences*)

Finding Nuggets in IP Portfolios: Core Patent Mining Through Textual Temporal Analysis (Page 1819)

Po Hu (*Tsinghua University*)
 Minlie Huang (*Tsinghua University*)
 Peng Xu (*ExxonMobil Research and Engineering Company*)
 Weichang Li (*ExxonMobil Research and Engineering Company*)
 Adam K. Usadi (*ExxonMobil Research and Engineering Company*)
 Xiaoyan Zhu (*Tsinghua University*)

Interest-Matching Information Propagation in Multiple Online Social Networks (Page 1824)

Yilin Shen (*University of Florida*)
 Thang N. Dinh (*University of Florida*)
 Huiyuan Zhang (*University of Florida*)
 My T. Thai (*University of Florida*)

Customizing Search Results for Non-Native Speakers (Page 1829)

Theodoros Lappas (*Boston University*)
 Michail Vlachos (*IBM Research*)

Quality Models for Microblog Retrieval (Page 1834)

Jaeho Choi (*NHN Corporation*)
 W. Bruce Croft (*University of Massachusetts, Amherst*)
 Jin Young Kim (*University of Massachusetts, Amherst*)

(Return to Top)

Do Ads Compete or Collaborate? Designing Click Models with Full Relationship Incorporated (Page 1839)

Xin Xin (*Beijing Institute of Technology*)
 Irwin King (*The Chinese University of Hong Kong*)
 Ritesh Agrawal (*AT&T Labs-Research*)
 Michael R. Lyu (*The Chinese University of Hong Kong*)
 Heyan Huang (*Beijing Institute of Technology*)

Exploiting Concept Hierarchy for Result Diversification (Page 1844)

Wei Zheng (*University of Delaware*)
 Hui Fang (*University of Delaware*)

Conglei Yao (*Tencent*)

Ranking News Events by Influence Decay and Information Fusion for Media and Users (Page 1849)

Liang Kong (*Peking University*)
 Shan Jiang (*Peking University*)
 Rui Yan (*Peking University*)
 Shize Xu (*Peking University*)
 Yan Zhang (*Peking University*)

Leveraging Tagging for Neighborhood-aware Probabilistic Matrix Factorization (Page 1854)

Le Wu (*University of Science and Technology of China*)
 Enhong Chen (*University of Science and Technology of China*)
 Qi Liu (*University of Science and Technology of China*)
 Linli Xu (*University of Science and Technology of China*)
 Tengfei Bao (*University of Science and Technology of China*)
 Lei Zhang (*University of Science and Technology of China*)

Semantic Context Learning with Large-Scale Weakly-Labeled Image Set (Page 1859)

Yao Lu (*Fudan University*)
 Wei Zhang (*Fudan University*)
 Ke Zhang (*Fudan University*)
 Xiangyang Xue (*Fudan University*)

Sketch-based Indexing of n-Words (Page 1864)

Samuel Huston (*University of Massachusetts, Amherst*)
 J. Shane Culpepper (*RMIT University*)
 W. Bruce Croft (*University of Massachusetts, Amherst*)

(Return to Top)

Interactive and Context-aware Tag Spell Check and Correction (Page 1869)

Francesco Bonchi (*Yahoo! Research*)
 Ophir Frieder (*Georgetown University*)
 Franco Maria Nardini (*ISTI-CNR*)
 Fabrizio Silvestri (*ISTI-CNR*)
 Hossein Vahabi (*ISTI-CNR*)

Federated Search in the Wild: The Combined Power of over a Hundred Search Engines (Page 1874)

Dong Nguyen (*University of Twente*)
 Thomas Demeester (*Ghent University*)
 Dolf Trieschnigg (*University of Twente*)
 Djoerd Hiemstra (*University of Twente*)

From sBoW to dCoT Marginalized Encoders for Text Representation (Page 1879)

Zhixiang (Eddie) Xu (*Washington University in St. Louis*)
 Minmin Chen (*Washington University in St. Louis*)
 Kilian Q. Weinberger (*Washington University in St. Louis*)

Fei Sha (*University of Southern California*)

Task Tours: Helping Users Tackle Complex Search Tasks (Page 1885)

Ahmed Hassan (*Microsoft Research*)

Ryen W. White (*Microsoft Research*)

Structured Query Reformulations in Commerce Search (Page 1890)

Sreenivas Gollapudi (*Microsoft Research Silicon Valley*)

Samuel Jeong (*Microsoft Research Silicon Valley*)

Anitha Kannan (*Microsoft Research Silicon Valley*)

Towards Jointly Extracting Aspects and Aspect-Specific Sentiment Knowledge (Page 1895)

Xueke Xu (*Chinese Academy of Sciences*)

Songbo Tan (*Chinese Academy of Sciences*)

Yue Liu (*Chinese Academy of Sciences*)

Xueqi Cheng (*Chinese Academy of Sciences*)

Zheng Lin (*Chinese Academy of Sciences*)

Collaborative Ranking: Improving the Relevance for Tail Queries (Page 1900)

Ke Zhou (*Georgia Institute of Technology*)

Xin Li (*Microsoft*)

Hongyuan Zha (*Georgia Institute of Technology*)

BiasTrust: Teaching Biased Users about Controversial Topics (Page 1905)

V. G. Vinod Vydiswaran (*University of Illinois*)

ChengXiang Zhai (*University of Illinois*)

Dan Roth (*University of Illinois*)

Peter Pirolli (*Palo Alto Research Center*)

Recommending Citations: Translating Papers into References (Page 1910)

Wenyi Huang (*The Pennsylvania State University*)

Saurabh Kataria (*Xerox Research Center Webster*)

Cornelia Caragea (*The Pennsylvania State University*)

Prasenjit Mitra (*The Pennsylvania State University*)

C. Lee Giles (*The Pennsylvania State University*)

Lior Rokach (*Ben-Gurion University of the Negev*)

Query-Biased Learning to Rank for Real-Time Twitter Search (Page 1915)

Xin Zhang (*Graduate University of Chinese Academy of Sciences*)

Ben He (*Graduate University of Chinese Academy of Sciences*)

Tiejian Luo (*Graduate University of Chinese Academy of Sciences*)

Baobin Li (*Graduate University of Chinese Academy of Sciences*)

Discovering Logical Knowledge for Deep Question Answering (Page 1920)

Zhao Liu (*Fudan University*)

Xipeng Qiu (*Fudan University*)

Ling Cao (*Fudan University*)

Xuanjing Huang (*Fudan University*)

Mining Noisy Tagging from Multi-Label Space (Page 1925)

Zhongang Qi (*Zhejiang University*)

Ming Yang (*Zhejiang University*)

Zhongfei (Mark) Zhang (*Zhejiang University*)

Zhengyou Zhang (*Microsoft Research*)

Learning from Mistakes: Towards a Correctable Learning Algorithm (Page 1930)

Karthik Raman (*Cornell University*)

Krysta M. Svore (*Microsoft Research*)

Ran Gilad-Bachrach (*Microsoft Research*)

Chris J. C. Burges (*Microsoft Research*)

CONSENTO: A New Framework for Opinion Based Entity Search and Summarization (Page 1935)

Jaehoon Choi (*Korea University*)

Donghyeon Kim (*Korea University*)

Seongsoon Kim (*Korea University*)

Junkyu Lee (*Korea University*)

Sangrak Lim (*Korea University*)

Sunwon Lee (*Korea University*)

Jaewoo Kang (*Korea University*)

Search Result Presentation Based on Faceted Clustering (Page 1940)

Benno Stein (*Bauhaus-Universität*)

Tim Gollub (*Bauhaus-Universität*)

Dennis Hoppe (*Bauhaus-Universität*)

PolariCQ: Polarity Classification of Political Quotations (Page 1945)

Rawia Awadallah (*Max-Planck Institute for Informatics*)

Maya Ramanath (*Indian Institute of Technology*)

Gerhard Weikum (*Max-Planck Institute for Informatics*)

A Comprehensive Analysis of Parameter Settings for Novelty-Biased Cumulative Gain (Page 1950)

Teerapong Leelanupab (*King Mongkut's Institute of Technology Ladkrabang*)

[\(Return to Top\)](#)

[\(Return to Top\)](#)

Guido Zuccon (*Australian e-Health Research Centre*)

Joemon M. Jose (*University of Glasgow*)

Entity Centric Query Expansion for Enterprise Search (Page 1955)

Xitong Liu (*University of Delaware*)

Hui Fang (*University of Delaware*)

Fei Chen (*HP Labs*)

Min Wang (*HP Labs*)

Location-sensitive Resources Recommendation in Social Tagging Systems (Page 1960)

Chang Wan (*The University of Hong Kong*)

Ben Kao (*The University of Hong Kong*)

David W. Cheung (*The University of Hong Kong*)

Differences in Effectiveness across Sub-Collections (Page 1965)

Mark Sanderson (*RMIT University*)

Andrew Turpin (*University of Melbourne*)

Ying Zhang (*RMIT University*)

Falk Scholer (*RMIT University*)

Map to Humans and Reduce Error - Crowdsourcing for Deduplication Applied to Digital Libraries (Page 1970)

Mihai Georgescu (*University of Hannover*)

Dang Duc Pham (*University of Hannover*)

Claudiu S. Firan (*University of Hannover*)

Wolfgang Nejdl (*University of Hannover*)

Julien Gaugaz (*University of Hannover*)

Full-Text Citation Analysis: Enhancing Bibliometric and Scientific Publication Ranking (Page 1975)

Xiaozhong Liu (*Indiana University, Bloomington*)

Jinsong Zhang (*Dalian Maritime University*)

Chun Guo (*Indiana University, Bloomington*)

Detecting Offensive Tweets via Topical Feature Discovery over a Large Scale Twitter Corpus (Page 1980)

Guang Xiang (*Carnegie Mellon University*)

Bin Fan (*Carnegie Mellon University*)

Ling Wang (*Carnegie Mellon University*)

Jason I. Hong (*Carnegie Mellon University*)

Carolyn P. Rose (*Carnegie Mellon University*)

Automatic Query Expansion Based on Tag Recommendation (Page 1985)

Vitor Oliveira (*Universidade Federal de Minas Gerais & Zunnit Technologies*)

Guilherme Gomes (*Universidade Federal de Minas Gerais*)

Fabiano Belém (*Universidade Federal de Minas Gerais*)

Wladimir Brandão (*Universidade Federal de Minas Gerais*)

Jussara Almeida (*Universidade Federal de Minas Gerais*)

Nivio Ziviani (*Universidade Federal de Minas Gerais & Zunnit Technologies*)

Marcos Gonçalves (*Universidade Federal de Minas Gerais*)

The Downside of Markup: Examining the Harmful Effects of CSS and Javascript on Indexing Today's Web (Page 1990)

Karl Gyllstrom (*Katholieke Universiteit Leuven*)

Carsten Eickhoff (*Delft University of Technology*)

Arjen P. de Vries (*Centrum Wiskunde & Informatica*)

Marie-Francine Moens (*Katholieke Universiteit Leuven*)

You Should Read This! Let Me Explain You Why: Explaining News Recommendations to Users (Page 1995)

Roi Blanco (*Yahoo! Research*)

Diego Ceccarelli (*ISTI-CNR & Università di Pisa*)

Claudio Lucchese (*ISTI-CNR & Università di Pisa*)

Raffaele Perego (*ISTI-CNR & Università di Pisa*)

Fabrizio Silvestri (*ISTI-CNR & Università di Pisa*)

Characterizing Web Search Queries That Match Very Few or No Results (Page 2000)

Ismail Sengor Altingovde (*University of Hannover*)

Roi Blanco (*Yahoo! Research*)

Berkant Barla Cambazoglu (*Yahoo! Research*)

Rifat Ozcan (*Bilkent University*)

Erdem Sarigil (*Bilkent University*)

Özgür Ulusoy (*Bilkent University*)

A Unified Optimization Framework for Auction and Guaranteed Delivery in Online Advertising (Page 2005)

Konstantin Salomatin (*Carnegie Mellon University*)

Tie-Yan Liu (*Microsoft Research Asia*)

Yiming Yang (*Carnegie Mellon University*)

Query Recommendation for Children (Page 2010)

Sergio Duarte Torres (*University of Twente*)

Djoerd Hiemstra (*University of Twente*)

Ingmar Weber (*Yahoo! Research*)

Pavel Serdyukov (*Yandex*)

Modeling Browsing Behavior for Click Analysis in Sponsored Search (Page 2015)

(Return to Top)

(Return to Top)

Azin Ashkan (*University of Waterloo*)
Charles L. A. Clarke (*University of Waterloo*)

Sentiment-Focused Web Crawling (Page 2020)

A. Gural Vural (*Middle East Technical University*)
B. Barla Cambazoglu (*Yahoo! Research*)
Pinar Senkul (*Middle East Technical University*)

User Guided Entity Similarity Search Using Meta-Path Selection in Heterogeneous Information Networks (Page 2025)

Xiao Yu (*University of Illinois at Urbana-Champaign*)
Yizhou Sun (*University of Illinois at Urbana-Champaign*)
Brandon Norick (*University of Illinois at Urbana-Champaign*)
Tiancheng Mao (*University of Illinois at Urbana-Champaign*)
Jiawei Han (*University of Illinois at Urbana-Champaign*)

User Activity Profiling with Multi-Layer Analysis (Page 2030)

Hongxia Jin (*Samsung Information System of America*)

GTE: A Distributional Second-Order Co-Occurrence Approach to Improve the Identification of Top Relevant Dates in Web Snippets (Page 2035)

Ricardo Campos (*LIAAD - INESC TEC, polytechnic Institute of Tomar & University of Beira Interior*)
Gaël Dias (*University of Caen Basse-Normandie & University of Beira Interior*)
Alípio Mário Jorge (*LIAAD - INESC TEC & University of Porto*)
Célia Nunes (*University of Beira Interior*)

Stochastic Simulation of Time-Biased Gain (Page 2040)

Mark D. Smucker (*University of Waterloo*)
Charles L. A. Clarke (*University of Waterloo*)

SonetRank: Leveraging Social Networks to Personalize Search (Page 2045)

Abhijith Kashyap (*University of California, Riverside*)
Reza Amini (*Florida International University*)
Vagelis Hristidis (*University of California, Riverside*)

Predicting Web Search Success with Fine-Grained Interaction Data (Page 2050)

Qi Guo (*Emory University*)
Dmitry Lagun (*Emory University*)
Eugene Agichtein (*Emory University*)

Multi-Session Re-Search: In Pursuit of Repetition and Diversification (Page 2055)

Sarah K. Tyler (*University of California, Santa Cruz*)
Yi Zhang (*University of California, Santa Cruz*)

Mining Sentiment Terminology Through Time (Page 2060)

Hadi Amiri (*National University of Singapore*)
Tat-Seng Chua (*National University of Singapore*)

Theme Chronicle Model: Chronicle Consists of Timestamp and Topical Words over Each Theme (Page 2065)

Noriaki Kawamae (*Tokyo Denki University*)

Fast Top-K Similarity Queries via Matrix Compression (Page 2070)

Yucheng Low (*Carnegie Mellon University*)

Alice X. Zheng (*Microsoft Research*)

Databases Short Paper Session

Top-k Retrieval Using Conditional Preference Networks (Page 2075)

Hongbing Wang (*Southeast University*)
Xuan Zhou (*Renmin University of China*)
Wujin Chen (*Southeast University*)
Peisheng Ma (*Southeast University*)

Sort-Based Query-adaptive Loading of R-Trees (Page 2080)

Daniar Achakeev (*Philipps-Universität Marburg*)
Bernhard Seeger (*Philipps-Universität Marburg*)
Peter Widmayer (*ETH Zürich*)

Efficient Logging for Enterprise Workloads on Column-Oriented In-Memory Databases (Page 2085)

Johannes Wust (*Hasso Plattner Institute*)
Joos-Hendrick Boese (*SAP AG*)
Frank Renkes (*SAP AG*)
Sebastian Blessing (*Hasso Plattner Institute*)
Jens Krueger (*Hasso Plattner Institute*)
Hasso Plattner (*Hasso Plattner Institute*)

Schema-Free Structured Querying of DBpedia Data (Page 2090)

Lushan Han (*University of Maryland, Baltimore County*)
Tim Finin (*University of Maryland, Baltimore County*)
Anupam Joshi (*University of Maryland, Baltimore County*)

Discovering Conditional Inclusion Dependencies (Page 2094)

Jana Bauckmann (*Hasso-Plattner-Institut*)

(Return to Top)

(Return to Top)

Ziawasch Abedjan (*Hasso-Plattner-Institut*)
 Ulf Leser (*Humboldt-Universität zu Berlin*)
 Heiko Müller (*Intelligent Sensing and Systems Laboratory*)
 Felix Naumann (*Hasso-Plattner-Institut*)

Diversifying Query Results on Semi-Structured Data (Page 2099)

Mahbub Hasan (*University of California, Riverside*)
 Abdullah Mueen (*University of California, Riverside*)
 Vassilis Tsotras (*University of California, Riverside*)
 Eamonn Keogh (*University of California, Riverside*)

LINDA: Distributed Web-of-Data-Scale Entity Matching (Page 2104)

Christoph Böhm (*Hasso Plattner Institute*)
 Gerard de Melo (*ICSI Berkeley*)
 Felix Naumann (*Hasso Plattner Institute*)
 Gerhard Weikum (*Max Planck Institute for Informatics*)

SliceSort: Efficient Sorting of Hierarchical Data (Page 2109)

Quoc Trung Tran (*University of California, Santa Cruz*)
 Chee-Yong Chan (*National University of Singapore*)

(Return to Top)

Efficient Buffer Management for Piecewise Linear Representation of Multiple Data Streams (Page 2114)

Qing Xie (*The University of Queensland*)
 Jia Zhu (*The University of Queensland*)
 Mohamed A. Sharaf (*The University of Queensland*)
 Xiaofang Zhou (*The University of Queensland*)
 Chaoyi Pang (*CSIRO*)

On Skyline Groups (Page 2119)

Chengkai Li (*University Texas at Arlington*)
 Nan Zhang (*George Washington University*)
 Naeemul Hassan (*University Texas at Arlington*)
 Sundaresan Rajasekaran (*George Washington University*)
 Gautam Das (*University Texas at Arlington, Qatar Computing Research Institute*)

Finding the Optimal Path over Multi-Cost Graphs (Page 2124)

Yajun Yang (*Harbin Institute of Technology*)
 Jeffrey Xu Yu (*The Chinese University of Hong Kong*)
 Hong Gao (*Harbin Institute of Technology*)
 Jianzhong Li (*Harbin Institute of Technology*)

An Efficient Index for Massive IOT Data in Cloud Environment (Page 2129)

Youzhong Ma (*Renmin University of China & Zhengzhou Chenggong University of Finance and Economics*)
 Jia Rao (*NEC Labs, China*)
 Weisong Hu (*NEC Labs, China*)
 Xiaofeng Meng (*Renmin University of China*)
 Xu Han (*Renmin University of China*)
 Yu Zhang (*Renmin University of China*)
 Yunpeng Chai (*Renmin University of China*)
 Chunqiu Liu (*Renmin University of China*)

Clustering Wikipedia Infoboxes to Discover Their Types (Page 2134)

Thanh Nguyen (*University of Utah*)
 Huong Nguyen (*University of Utah*)
 Viviane Moreira (*UFRGS-Brazil*)
 Juliana Freire (*NYU Poly*)

CloST: A Hadoop-based Storage System for Big Spatio-Temporal Data Analytics (Page 2139)

Haoyu Tan (*Hong Kong University of Science and Technology*)
 Wuman Luo (*Hong Kong University of Science and Technology*)
 Lionel M. Ni (*Hong Kong University of Science and Technology*)

Keyword-based k-Nearest Neighbor Search in Spatial Databases (Page 2144)

Guoliang Li (*Tsinghua University*)
 Jing Xu (*Tsinghua University*)
 Jianhua Feng (*Tsinghua University*)

Credibility-Based Product Ranking for C2C Transactions (Page 2149)

Rong Zhang (*East China Normal University*)
 ChaoFeng Sha (*Fudan University*)
 Minqi Zhou (*East China Normal University*)
 Ao Ying Zhou (*East China Normal University*)

(Return to Top)

Location Selection for Utility Maximization with Capacity Constraints (Page 2154)

Yu Sun (*Renmin University of China*)
 Jin Huang (*University of Melbourne*)
 Yueguo Chen (*Renmin University of China*)
 Rui Zhang (*University of Melbourne*)
 Xiaoyong Du (*Renmin University of China*)

Efficient Estimation of Dynamic Density Functions with an Application to Outlier Detection (Page 2159)

Abdulkhakim Ali Qahtan (*King Abdullah University of Sciences & Technology*)

Xiangliang Zhang (*King Abdullah University of Sciences & Technology*)
 Suojin Wang (*Texas A&M University*)

A Positional Access Method for Relational Databases (Page 2164)

Dongzhe Ma (*Tsinghua University*)
 Jianhua Feng (*Tsinghua University*)
 Guoliang Li (*Tsinghua University*)

Real-Time Aggregate Monitoring with Differential Privacy (Page 2169)

Liyue Fan (*Emory University*)
 Li Xiong (*Emory University*)

Efficient Distributed Locality Sensitive Hashing (Page 2174)

Bahman Bahmani (*Stanford University*)
 Ashish Goel (*Stanford University*)
 Rajendra Shinde (*Stanford University*)

Author-Conference Topic-Connection Model for Academic Network Search (Page 2179)

Jianwen Wang (*Central China Normal University*)
 Xiaohua Hu (*Central China Normal University & Drexel University*)
 Xinhui Tu (*Central China Normal University*)
 Tingting He (*Central China Normal University*)

Impact Neighborhood Indexing (INI) in Diffusion Graphs (Page 2184)

Jung Hyun Kim (*Arizona State University*)
 K. Selçuk Candan (*Arizona State University*)
 Maria Luisa Sapino (*University of Torino*)

Loyalty-based Selection: Retrieving Objects that Persistently Satisfy Criteria (Page 2189)

Zhitao Shen (*The University of New South Wales*)
 Muhammad Aamir Cheema (*The University of New South Wales*)
 Xuemin Lin (*The University of New South Wales*)

Star-Join: Spatio-Textual Similarity Join (Page 2194)

Sitong Liu (*Tsinghua University*)
 Guoliang Li (*Tsinghua University*)
 Jianhua Feng (*Tsinghua University*)

Adapt: Adaptive Database Schema Design for Multi-Tenant Applications (Page 2199)

Jiacai Ni (*Tsinghua University*)
 Guoliang Li (*Tsinghua University*)
 Jun Zhang (*IBM China Development Lab*)
 Lei Li (*IBM China Development Lab*)
 Jianhua Feng (*Tsinghua University*)

Optimizing Data Migration for Cloud-Based Key-Value Stores (Page 2204)

Xiulei Qin (*Chinese Academy of Sciences*)
 Wenbo Zhang (*Chinese Academy of Sciences*)
 Wei Wang (*Chinese Academy of Sciences*)
 Jun Wei (*Chinese Academy of Sciences*)
 Xin Zhao (*Chinese Academy of Sciences*)
 Tao Huang (*Chinese Academy of Sciences*)

(Return to Top)

Applying Weighted Queries on Probabilistic Databases (Page 2209)

Sebastian Lehrack (*Brandenburg University of Technology Cottbus*)

A New Tool for Multi-Level Partitioning in Teradata (Page 2214)

Young-Kyoon Suh (*University of Arizona*)
 Ahmad Ghazal (*Teradata Corporation*)
 Alain Crolotte (*Teradata Corporation*)
 Pekka Kostamaa (*Teradata Corporation*)

Fast PCA Computation in a DBMs with Aggregate UDFs and LAPACK (Page 2219)

Carlos Ordonez (*University of Houston*)
 Naveen Mohanam (*University of Houston*)
 Carlos Garcia-Alvarado (*University of Houston*)
 Predrag T. Tosic (*University of Houston*)
 Edgar Martinez (*University of Houston*)

Scaling Multiple-Source Entity Resolution Using Statistically Efficient Transfer Learning (Page 2224)

Sahand Negahban (*Massachusetts Institute of Technology*)
 Benjamin I. P. Rubinstein (*Microsoft Research*)
 Jim Gemmell (*Microsoft Research*)

A Probabilistic Approach to Correlation Queries in Uncertain Time Series Data (Page 2229)

Mahsa Orang (*Concordia University*)
 Nematollah Shiri (*Concordia University*)

On Bundle Configuration for Viral Marketing in Social Networks (Page 2234)

De-Nian Yang (*Academia Sinica*)
 Wang-Chien Lee (*The Pennsylvania State University*)
 Nai-Hui Chia (*Academia Sinica*)
 Mao Ye (*The Pennsylvania State University*)

Hui-Ju Hung (*Academia Sinica*)**Knowledge Management Poster Session****Learning to Rank for Hybrid Recommendation** (Page 2239)Jiankai Sun (*Shandong University*)Shuaiqiang Wang (*Shandong University of Finance and Economics*)Byron J. Gao (*Texas State University at San Marcos*)Jun Ma (*Shandong University*)**Importance Weighted Passive Learning** (Page 2243)Shuaiqiang Wang (*Shandong University of Finance and Economics*)Xiaoming Xi (*Shandong University*)Yilong Yin (*Shandong University*)[\(Return to Top\)](#)**A Tag-Centric Discriminative Model for Web Objects Classification** (Page 2247)Lina Yao (*The University of Adelaide*)Quan Z. Sheng (*The University of Adelaide*)**Outlier Detection Using Centrality and Center-Proximity** (Page 2251)Duck-Ho Bae (*Hanyang University*)Seo Jeong (*Hanyang University*)Sang-Wook Kim (*Hanyang University*)Minsoo Lee (*Ewha Womans University*)**An Effective Category Classification Method Based on a Language Model for Question Category Recommendation on a cQA Service** (Page 2255)Kyoungman Bae (*Dong-A University*)Youngjoong Ko (*Dong-A University*)**Clustering Short Text Using Ncut-Weighted Non-Negative Matrix Factorization** (Page 2259)Xiaohui Yan (*Chinese Academy of Sciences*)Jiafeng Guo (*Chinese Academy of Sciences*)Shenghua Liu (*Chinese Academy of Sciences*)Xueqi Cheng (*Chinese Academy of Sciences*)Yanfeng Wang (*Sogou Inc.*)**Polygene-Based Evolution: A Novel Framework for Evolutionary Algorithms** (Page 2263)Shuaiqiang Wang (*Shandong University of Finance and Economics & Texas State University-San Marcos*)Byron J. Gao (*Texas State University-San Marcos*)Shuangling Wang (*Shandong University*)Guibao Cao (*Shandong University*)Yilong Yin (*Shandong University*)**A Tensor Encoding Model for Semantic Processing** (Page 2267)Michael Symonds (*Queensland University of Technology*)Peter Bruza (*Queensland University of Technology*)Laurianne Sitbon (*Queensland University of Technology*)Ian Turner (*Queensland University of Technology*)[\(Return to Top\)](#)**Accelerating Locality Preserving Nonnegative Matrix Factorization** (Page 2271)Guanhong Yao (*Zhejiang University*)Deng Cai (*Zhejiang University*)**The Twitaholic Next Door: Scalable friend recommender system using a concept-sensitive hash function** (Page 2275)Patrick Bamba (*Université Jean Monnet*)Julien Subercaze (*Université Jean Monnet*)Christophe Gravier (*Université Jean Monnet*)Nabil Benmira (*Graphinium - ESSEC VENTURES*)Jimi Fontaine (*Graphinium - ESSEC VENTURES*)**Information Propagation in Social Rating Networks** (Page 2279)Priyanka Garg (*The Chinese University of Hong Kong*)Irwin King (*The Chinese University of Hong Kong*)Michael R. Lyu (*The Chinese University of Hong Kong*)**Maximizing Revenue from Strategic Recommendations under Decaying Trust** (Page 2283)Paul Dütting (*École Polytechnique Fédérale de Lausanne*)Monika Henzinger (*University of Vienna*)Ingmar Weber (*Yahoo! Research*)**Weighted Linear Kernel with Tree Transformed Features for Malware Detection** (Page 2287)Prakash Mandayam Comar (*Michigan State University*)Lei Liu (*Michigan State University*)Sabyasachi Saha (*Narus Inc.*)Pang-Ning Tan (*Michigan State University*)Antonio Nucci (*Narus Inc.*)**Learning to Predict the Cost-Per-Click for Your Ad Words** (Page 2291)Chieh-Jen Wang (*National Taiwan University*)Hsin-Hsi Chen (*National Taiwan University*)[\(Return to Top\)](#)**Dual Word and Document Seed Selection for Semi-supervised Sentiment Classification** (Page 2295)Shih-Wei Fan (*Google Inc.*)Shih-Wei Fan (*Google Inc.*)

Shengteng Ju (*Soochow University*)
 Shoushan Li (*Soochow University*)
 Yan Su (*Soochow University*)
 Guodong Zhou (*Soochow University*)
 Yu Hong (*Soochow University*)
 Xiaojun Li (*Zhejiang Gongshang University*)

On Empirical Tradeoffs in Large Scale Hierarchical Classification (Page 2299)

Rohit Babbar (*Joseph Fourier University*)
 Ioannis Partalas (*Joseph Fourier University*)
 Eric Gaussier (*Joseph Fourier University*)
 Cecile Amblard (*Joseph Fourier University*)

An Interaction Framework of Service-oriented Ontology Learning (Page 2303)

Jingsong Zhang (*Shanghai Jiao Tong University*)
 Yinglin Wang (*Shanghai Jiao Tong University*)
 Hao Wei (*University of Fribourg*)

Infobox Suggestion for Wikipedia Entities (Page 2307)

Afroza Sultana (*University of Texas at Arlington*)
 Quazi M. Hasan (*Dematic*)
 Ashis K. Biswas (*University of Texas at Arlington*)
 Soumyava Das (*University of Texas at Arlington*)
 Habibur Rahman (*University of Texas at Arlington*)
 Chris Ding (*University of Texas at Arlington*)
 Chengkai Li (*University of Texas at Arlington*)

Time Feature Selection for Identifying Active Household Members (Page 2311)

Pedro G. Campos (*Universidad Autónoma de Madrid & Universidad del Bio-Bio*)
 Alejandro Bellogin (*Universidad Autónoma de Madrid*)
 Fernando Diez (*Universidad Autónoma de Madrid*)
 Iván Cantador (*Universidad Autónoma de Madrid*)

Text Classification with Relatively Small Positive Documents and Unlabeled Data (Page 2315)

Fumiyo Fukumoto (*University of Yamanashi*)
 Takeshi Yamamoto (*University of Yamanashi*)
 Suguru Matsuyoshi (*University of Yamanashi*)
 Yoshimi Suzuki (*University of Yamanashi*)

On Compressing Weighted Time-evolving Graphs (Page 2319)

Wei Liu (*The University of Melbourne*)
 Andrey Kan (*The University of Melbourne*)
 Jeffrey Chan (*The University of Melbourne*)
 James Bailey (*The University of Melbourne*)
 Christopher Leckie (*The University of Melbourne*)
 Pei Jian (*Simon Fraser University*)
 Ramamohanarao Kotagiri (*The University of Melbourne*)

Graph-based Collective Classification for Tweets (Page 2323)

Yajuan Duan (*University of Science and Technology of China*)
 Furu Wei (*Microsoft Research Asia*)
 Ming Zhou (*Microsoft Research Asia*)
 Heung-Yeung Shum (*Microsoft Corporation*)

A Word-Order Based Graph Representation for Relevance Identification (Page 2327)

Lakshmi Ramachandran (*North Carolina State University*)
 Edward F. Gehringer (*North Carolina State University*)

Tracing Clusters in Evolving Graphs with Node Attributes (Page 2331)

Brigitte Boden (*RWTH Aachen University*)
 Stephan Günnemann (*RWTH Aachen University*)
 Thomas Seidl (*RWTH Aachen University*)

Prediction of Retweet Cascade Size over Time (Page 2335)

Andrey Kupavskii (*Yandex Corporate*)
 Liudmila Ostroumova (*Yandex Corporate*)
 Alexey Umnov (*Yandex Corporate*)
 Svyatoslav Usachev (*Yandex Corporate*)
 Pavel Serdyukov (*Yandex Corporate*)
 Gleb Gusev (*Yandex Corporate*)
 Andrey Kustarev (*Yandex Corporate*)

An Efficient and Simple Under-sampling Technique for Imbalanced Time Series Classification (Page 2339)

Guohua Liang (*University of Technology Sydney*)
 Chengqi Zhang (*University of Technology Sydney*)

Top-N Recommendation through Belief Propagation (Page 2343)

Jiwoon Ha (*Hanyang University*)
 Soon-Hyoung Kwon (*Hanyang University*)
 Sang-Wook Kim (*Hanyang University*)
 Christos Faloutsos (*Carnegie Mellon University*)
 Sunju Park (*Yonsei University*)

(Return to Top)

Mining Advices from Weblogs (Page 2347)Alfan Farizki Wicaksono (*Korea Advanced Institute of Science and Technology*)Sung-Hyon Myaeng (*Korea Advanced Institute of Science and Technology*)**Parallel Proximal Support Vector Machine for High-Dimensional Pattern Classification** (Page 2351)Zhenfeng Zhu (*Zhengzhou University*)Xingquan Zhu (*University of Technology Sydney*)Yangdong Ye (*Zhengzhou University*)Yue-Fei Guo (*Fudan University*)Xiangyang Xue (*Fudan University*)**On Using Category Experts for Improving the Performance and Accuracy in Recommender Systems** (Page 2355)Won-Seok Hwang (*Hanyang University*)Ho-Jong Lee (*Hanyang University*)Sang-Wook Kim (*Hanyang University*)Minsoo Lee (*Ewha Womans University*)**Finding Influential Products on Social Domination Game** (Page 2359)Jinyoung Yeo (*Pohang University of Science and Technology*)Jin-woo Park (*Pohang University of Science and Technology*)Seung-won Hwang (*Pohang University of Science and Technology*)**Entity Resolution using Search Engine Results** (Page 2363)Madian Khabsa (*The Pennsylvania State University*)Pucktada Treeratpituk (*The Pennsylvania State University*)C. Lee Giles (*The Pennsylvania State University*)**Tweet Classification Based on Their Lifetime Duration** (Page 2367)Hikaru Takemura (*Kyoto University*)Keishi Tajima (*Kyoto University*)**Scalable Collaborative Filtering Using Incremental Update and Local Link Prediction** (Page 2371)Xiao Yang (*Harbin Institute of Technology*)Zhaoxin Zhang (*Harbin Institute of Technology*)Ke Wang (*Harbin Institute of Technology*)**Composing Activity Groups in Social Networks** (Page 2375)Cheng-Te Li (*National Taiwan University*)Man-Kwan Shan (*National Chengchi University*)**A Co-training based Method for Chinese Patent Semantic Annotation** (Page 2379)Xu Chen (*Wuhan University*)Zhiyong Peng (*Wuhan University*)Cheng Zeng (*Wuhan University*)**Automatic Labeling Hierarchical Topics** (Page 2383)Xian-Ling Mao (*Peking University*)Zhao-Yan Ming (*National University of Singapore*)Zheng-Jun Zha (*National University of Singapore*)Tat-Seng Chua (*National University of Singapore*)Hongfei Yan (*Peking University*)Xiaoming Li (*Peking University*)**An Unsupervised Method for Author Extraction from Web Pages Containing User-Generated Content** (Page 2387)Jing Liu (*Harbin Institute of Technology*)Xinying Song (*Microsoft Research*)Jingtian Jiang (*Microsoft Research Asia*)Chin-Yew Lin (*Microsoft Research Asia*)**Hierarchical Target Type Identification for Entity-oriented Queries** (Page 2391)Krisztian Balog (*Norwegian University of Science and Technology*)Robert Neumayer (*Norwegian University of Science and Technology*)**Dictionary based Sparse Representation for Domain Adaptation** (Page 2395)Rishabh Mehrotra (*BITS Pilani*)Rushabh Agrawal (*BITS Pilani*)Syed Aqueel Haider (*MIT Manipal*)**Information Retrieval Poster Session****Selecting Expansion Terms as a Set via Integer Linear Programming** (Page 2399)Qi Zhang (*Fudan University*)Yan Wu (*Fudan University*)Xuanjing Huang (*Fudan University*)**An Evaluation and Enhancement of Densitometric Fragmentation for Content Slicing Reuse** (Page 2403)Killian Levacher (*Trinity College Dublin*)Seamus Lawless (*Trinity College Dublin*)Vincent Wade (*Trinity College Dublin*)**Mathematical Equation Retrieval Using Plain Words as a Query** (Page 2407)Shinil Kim (*Dong-A University*)[\(Return to Top\)](#)[\(Return to Top\)](#)[\(Return to Top\)](#)

Seon Yang (*Dong-A University*)
 Youngjoong Ko (*Dong-A University*)

Serial Position Effects of Clicking Behavior on Result Pages Returned by Search Engines (Page 2411)

Mingda Wu (*Peking University*)
 Shan Jiang (*Peking University*)
 Yan Zhang (*Peking University*)

Towards Measuring the Visualness of a Concept (Page 2415)

Jin-Woo Jeong (*Hanyang University*)
 Xin-Jing Wang (*Microsoft Research Asia*)
 Dong-Ho Lee (*Hanyang University*)

Fast Candidate Generation for Two-Phase Document Ranking: Postings List Intersection with Bloom Filters (Page 2419)

Nima Asadi (*University of Maryland*)

Jimmy Lin (*University of Maryland*)

Semantically Coherent Image Annotation with a Learning-Based Keyword Propagation Strategy (Page 2423)

Chaoran Cui (*Shandong University & Shandong Provincial Key Laboratory of Software Engineering*)
 Jun Ma (*Shandong University & Shandong Provincial Key Laboratory of Software Engineering*)
 Shuaiqiang Wang (*Shandong University of Finance and Economics*)
 Shuai Gao (*Shandong University & Shandong Provincial Key Laboratory of Software Engineering*)
 Tao Lian (*Shandong University & Shandong Provincial Key Laboratory of Software Engineering*)

Language Processing for Arabic Microblog Retrieval (Page 2427)

Kareem Darwish (*Qatar Computing Research Institute*)
 Walid Magdy (*Qatar Computing Research Institute*)
 Ahmed Mourad (*Qatar Computing Research Institute*)

Hierarchical Image Annotation Using Semantic Hierarchies (Page 2431)

Hichem Bannour (*École Centrale Paris*)
 Céline Hudelot (*École Centrale Paris*)

On the Inference of Average Precision from Score Distributions (Page 2435)

Ronan Cummins (*National University of Ireland, Galway*)

An Evaluation of Corpus-Driven Measures of Medical Concept Similarity for Information Retrieval (Page 2439)

Bevan Koopman (*CSIRO & Queensland University of Technology*)
 Guido Zuccon (*CSIRO & Queensland University of Technology*)
 Peter Bruza (*Queensland University of Technology*)
 Laurianne Sitbon (*Queensland University of Technology*)
 Michael Lawley (*CSIRO*)

A Constraint to Automatically Regulate Document-Length Normalisation (Page 2443)

Ronan Cummins (*National University of Ireland Galway*)
 Colm O'Riordan (*National University of Ireland Galway*)

Bridging Offline and Online Social Graph Dynamics (Page 2447)

Manuel Gomez Rodriguez (*Stanford University*)
 Monica Rogati (*LinkedIn*)

Predicting the Performance of Passage Retrieval for Question Answering (Page 2451)

Eyal Krikon (*Technion*)
 David Carmel (*IBM Research - Haifa lab*)
 Oren Kurland (*Technion*)

Coarse-to-Fine Sentence-level Emotion Classification Based on the Intra-sentence Features and Sentential Context (Page 2455)

Jun Xu (*Harbin Institute of Technology*)
 Ruifeng Xu (*Harbin Institute of Technology*)
 Qin Lu (*The Hong Kong Polytechnic University*)
 Xiaolong Wang (*Harbin Institute of Technology*)

Query-Performance Prediction and Cluster Ranking: Two Sides of the Same Coin (Page 2459)

Oren Kurland (*Technion - Israel Institute of Technology*)
 Fiana Raiber (*Technion - Israel Institute of Technology*)
 Anna Shtok (*Technion - Israel Institute of Technology*)

Learning to Rank Search Results for Time-Sensitive Queries (Page 2463)

Nattiya Kanhabua (*Leibniz Universität Hannover*)
 Kjetil Nørkvåg (*Norwegian University of Science and Technology*)

On Active Learning in Hierarchical Classification (Page 2467)

Yu Cheng (*Northwestern University*)
 Kunpeng Zhang (*Northwestern University*)
 Yusheng Xie (*Northwestern University*)
 Ankit Agrawal (*Northwestern University*)
 Alok Choudhary (*Northwestern University*)

Question-Answer Topic Model for Question Retrieval in Community Question Answering (Page 2471)

Zongcheng Ji (*Chinese Academy of Sciences*)
 Fei Xu (*Chinese Academy of Sciences*)
 Bin Wang (*Chinese Academy of Sciences*)
 Ren He (*Graduate University of Chinese Academy of Sciences*)

[\(Return to Top\)](#)

[\(Return to Top\)](#)

Beihang University (Graduate University of Chinese Academy of Sciences)

How Do Humans Distinguish Different People with Identical Names on the Web?: A Cognitive Science Approach (Page 2475)

Harumi Murakami (*Osaka City University*)

Yuki Miyake (*Aeon Delight Co., Ltd.*)

Enhancing Product Search by Best-Selling Prediction in E-Commerce (Page 2479)

Bo Long (*Yahoo! Labs*)

Jiang Bian (*Yahoo! Labs*)

Anlei Dong (*Yahoo! Labs*)

Yi Chang (*Yahoo! Labs*)

Survival Analysis for Freshness in Microblogging Search (Page 2483)

Gianni Amati (*Fondazione Ugo Bordoni*)

Giuseppe Amodio (*Almawave*)

Carlo Gaibisso (*IASI-CNR*)

Information Preservation in Static Index Pruning (Page 2487)

Ruey-Cheng Chen (*National Taiwan University*)

Chia-Jung Lee (*University of Massachusetts*)

Chiung-Min Tsai (*National Taiwan University*)

Jieh Hsiang (*National Taiwan University*)

Temporal Models for Microblogs (Page 2491)

Jaeho Choi (*NHN Corporation*)

W. Bruce Croft (*University of Massachusetts, Amherst*)

I want what I need! Analyzing Subjectivity of Online Forum Threads (Page 2495)

Prakhar Biyani (*The Pennsylvania State University*)

Comelia Caragea (*The Pennsylvania State University*)

Amit Singh (*IBM Research*)

Prasenjit Mitra (*The Pennsylvania State University*)

Improving the Performance of the Reinforcement Learning Model for Answering Complex Questions (Page 2499)

Yllias Chali (*University of Lethbridge*)

Sadid A. Hasan (*University of Lethbridge*)

Kaisar Imam (*University of Lethbridge*)

Relation Regularized Subspace Recommending for Related Scientific Articles (Page 2503)

Qing Zhang (*Beijing Institute of Technology*)

Jianwu Li (*Beijing Institute of Technology*)

Zhiping Zhang (*Beijing Institute of Technology*)

Li Wang (*Beijing Institute of Technology*)

Exploring the Cluster Hypothesis, and Cluster-Based Retrieval, over the Web (Page 2507)

Fiana Raiber (*Technion - Israel Institute of Technology*)

Oren Kurland (*Technion - Israel Institute of Technology*)

A Picture Paints a Thousand Words: A Method of Generating Image-Text Timelines (Page 2511)

Shize Xu (*Peking University*)

Liang Kong (*Peking University*)

Yan Zhang (*Peking University*)

Short-Text Domain Specific Key Terms/Phrases Extraction Using an n-Gram Model with Wikipedia (Page 2515)

M. Atif Qureshi (*National University of Ireland, Galway & University of Milan Bicocca*)

Colm O'Riordan (*National University of Ireland Galway*)

Gabriella Pasi (*University of Milan Bicocca*)

A New Probabilistic Model for Top-k Ranking Problem (Page 2519)

Shuzi Niu (*Chinese Academy of Sciences*)

Yanyan Lan (*Chinese Academy of Sciences*)

Jiafeng Guo (*Chinese Academy of Sciences*)

Xueqi Cheng (*Chinese Academy of Sciences*)

Large Scale Analysis of Changes in English Vocabulary Over Recent Time (Page 2523)

Adam Jatowt (*Kyoto University & Japan Science and Technology Agency*)

Katsumi Tanaka (*Kyoto University*)

Climbing the App Wall: Enabling Mobile App Discovery through Context-Aware Recommendations (Page 2527)

Alexandros Karatzoglou (*Telefonica Research*)

Linas Baltrunas (*Telefonica Research*)

Karen Church (*Telefonica Research*)

Matthias Böhmer (*German Research Center for Artificial Intelligence*)

Twisent: A Multistage System for Analyzing Sentiment in Twitter (Page 2531)

Subhabrata Mukherjee (*IIT Bombay*)

Akshat Malu (*IIT Bombay*)

Balamurali A.R. (*IIT Bombay & IITB-Monash Research Academy*)

Pushpak Bhattacharyya (*IIT Bombay*)

Twitter Hyperlink Recommendation with User-Tweet-Hyperlink Three-way Clustering (Page 2535)

Dehong Gao (*The Hong Kong Polytechnic University*)

Renxian Zhang (*The Hong Kong Polytechnic University*)

Wenjie Li (*The Hong Kong Polytechnic University*)

(Return to Top)

(Return to Top)

Yuexian Hou (*Tianjin University*)

Concavity in IR Models (Page 2539)

Stéphane Clinchant (*Xerox Research Center Europe*)

Extracting Interesting Association Rules from Toolbar Data (Page 2543)

Ilaria Bordino (*Yahoo! Research*)

Debora Donato (*Yahoo! Labs*)

Barbara Pobleto (*Yahoo! Research*)

Predicting CTR of New Ads via Click Prediction (Page 2547)

Alexander Kolesnikov (*Yandex*)

Yury Logachev (*Yandex*)

Valeriy Topinskiy (*Yandex*)

An Examination of Content Farms in Web Search Using Crowdsourcing (Page 2551)

Richard McCreddie (*University of Glasgow*)

Craig Macdonald (*University of Glasgow*)

Iadh Ounis (*University of Glasgow*)

Jim Giles (*New Scientist*)

Ferris Jabr (*New Scientist*)

(Return to Top)

Demographic Context in Web Search Re-ranking (Page 2555)

Eugene Kharitonov (*Yandex, LLC & University of Glasgow*)

Pavel Serdyukov (*Yandex, LLC*)

On the Usefulness of Query Features for Learning to Rank (Page 2559)

Craig Macdonald (*University of Glasgow*)

Rodrygo L. T. Santos (*University of Glasgow*)

Iadh Ounis (*University of Glasgow*)

Session-based Query Performance Prediction (Page 2563)

Andrey Kustarev (*Yandex*)

Yury Ustinovskiy (*Yandex*)

Anna Mazur (*Yandex*)

Pavel Serdyukov (*Yandex*)

A Latent Pairwise Preference Learning Approach for Recommendation from Implicit Feedback (Page 2567)

Yi Fang (*Santa Clara University*)

Luo Si (*Purdue University*)

Topic Based Pose Relevance Learning in Dance Archives (Page 2571)

Reede Ren (*University of Glasgow*)

John Collomosse (*University of Surrey*)

Joemon Jose (*University of Glasgow*)

PhotoFall: Discovering Weblog Stories Through Photographs (Page 2575)

Christopher Wienberg (*University of Southern California*)

Andrew S. Gordon (*University of Southern California*)

RESQ: Rank-Energy Selective Query Forwarding for Distributed Search Systems (Page 2579)

Amin Teymorian (*Georgetown University*)

Xiao Qin (*Rutgers University*)

Ophir Frieder (*Georgetown University*)

The Face of Quality in Crowdsourcing Relevance Labels: Demographics, Personality and Labeling Accuracy (Page 2583)

Gabriella Kazai (*Microsoft Research*)

Jaap Kamps (*University of Amsterdam*)

Natasa Milic-Frayling (*Microsoft Research*)

Data Filtering in Humor Generation: Comparative Analysis of Hit Rate and Co-Occurrence Rankings as a Method to Choose Usable Pun Candidates (Page 2587)

Pawel Dybala (*Otaru University of Commerce*)

Rafal Rzepka (*Hokkaido University*)

Kenji Araki (*Hokkaido University*)

Kohichi Sayama (*Otaru University of Commerce*)

(Return to Top)

Predicting Primary Categories of Business Listings for Local Search (Page 2591)

Changsung Kang (*Yahoo! Labs*)

Jeehaeng Lee (*Yahoo! Labs*)

Yi Chang (*Yahoo! Labs*)

Where Do the Query Terms Come From? An Analysis of Query Reformulation in Collaborative Web Search (Page 2595)

Zhen Yue (*University of Pittsburgh*)

Jiepu Jiang (*University of Pittsburgh*)

Shuguang Han (*University of Pittsburgh*)

Daqing He (*University of Pittsburgh*)

Learning to Recommend with Social Relation Ensemble (Page 2599)

Lei Guo (*Shandong University & Shandong Provincial Key Laboratory of Software Engineering*)

Jun Ma (*Shandong University & Shandong Provincial Key Laboratory of Software Engineering*)

Zhumin Chen (*Shandong University & Shandong Provincial Key Laboratory of Software Engineering*)

Haoran Jiang (*Shandong Province Post Company*)

Albrecht Kothermel (*Ulm University*)

Continuous Top-k Query for Graph Streams (Page 2659)

Shirui Pan (*University of Technology, Sydney*)

Xingquan Zhu (*University of Technology, Sydney*)

(Return to Top)

Latent Topics in Graph-Structured Data (Page 2663)

Christoph Böhm (*Hasso Plattner Institute*)

Gjergji Kasneci (*Hasso Plattner Institute*)

Felix Naumann (*Hasso Plattner Institute*)

Fast and Accurate Incremental Entity Resolution Relative to an Entity Knowledge Base (Page 2667)

Michael Welch (*Barnes & Noble*)

Aamod Sane (*Yahoo! Inc.*)

Chris Drome (*Yahoo! Inc.*)

Knowledge Management Demonstration Session

LUKE and MIKE: Learning from User Knowledge and Managing Interactive Knowledge Extraction (Page 2671)

Steffen Metzger (*Max-Planck-Institute for Informatics*)

Michael Stoll (*University of Stuttgart*)

Katja Hose (*Max-Planck-Institute for Informatics*)

Ralf Schenkel (*Saarland University & Max-Planck-Institute for Informatics*)

PAVDA-live: Interactive Knowledge Harvesting (Page 2674)

Yafang Wang (*Max Planck Institute for Informatics*)

Maximilian Dylla (*Max Planck Institute for Informatics*)

Zhaochun Ren (*Max Planck Institute for Informatics*)

Marc Spaniol (*Max Planck Institute for Informatics*)

Gerhard Weikum (*Max Planck Institute for Informatics*)

(Return to Top)

4Is of Social Bully Filtering: Identity, Inference, Influence, and Intervention (Page 2677)

Yunfei Chen (*University of California, Santa Cruz*)

Lanbo Zhang (*University of California, Santa Cruz*)

Aaron Michelony (*University of California, Santa Cruz*)

Yi Zhang (*University of California, Santa Cruz*)

Ionomics Atlas - A Tool to Explore Interconnected Ionomic, Genomic and Environmental Data (Page 2680)

Eduard C. Dragut (*Purdue University*)

Mourad Ouzzani (*Qatar Computing Research Institute*)

Amgad Madkour (*Purdue University*)

Nabeel Mohamed (*Purdue University*)

Peter Baker (*Purdue University*)

David E. Salt (*University of Aberdeen*)

CarbonDB: A Semantic Life Cycle Inventory Database (Page 2683)

Benjamin Bertin (*Université de Lyon*)

Vasile-Marian Scuturici (*Université de Lyon*)

Jean-Marie Pinon (*Université de Lyon*)

Emmanuel Risler (*Université de Lyon*)

Supporting Temporal Analytics for Health-Related Events in Microblogs (Page 2686)

Nattiya Kanhabua (*University of Hannover*)

Avaré Stewart (*University of Hannover*)

Wolfgang Nejdl (*University of Hannover*)

Sara Romano (*Federico II University*)

InCaToMi: Integrative Causal Topic Miner Between Textual and Non-Textual Time Series Data (Page 2689)

Hyun Duk Kim (*University of Illinois at Urbana-Champaign*)

ChengXiang Zhai (*University of Illinois at Urbana-Champaign*)

Thomas A. Rietz (*The University of Iowa*)

Daniel Diermeier (*Northwestern University*)

Meichun Hsu (*HP Laboratories*)

Malu Castellanos (*HP Laboratories*)

Carlos A. Ceja Limon (*HP Laboratories*)

(Return to Top)

A Tool for Automated Evaluation of Algorithms (Page 2692)

Philipp Kranen (*RWTH Aachen University*)

Stephan Wels (*RWTH Aachen University*)

Tim Rohlfis (*RWTH Aachen University*)

Sebastian Raubach (*RWTH Aachen University*)

Thomas Seidl (*RWTH Aachen University*)

Information Retrieval Demonstration Session

A Summarization Tool for Time-Sensitive Social Media (Page 2695)

Walid Magdy (*Qatar Computing Research Institute*)

Ahmed Ali (*Qatar Computing Research Institute*)

Kareem Darwish (*Qatar Computing Research Institute*)

CrowdTiles: Presenting Crowd-based Information for Event-driven Information Needs (Page 2698)

Stewart Whiting (*University of Glasgow*)

Stewart Whiting (*University of Glasgow*)Ke Zhou (*University of Glasgow*)Joemon M. Jose (*University of Glasgow*)Omar Alonso (*Microsoft Corporation*)Teerapong Leelanupab (*King Mongkut's Institute of Technology*)[\(Return to Top\)](#)**ESA: Emergency Situation Awareness via Microbloggers** (Page 2701)Jie Yin (*CSIRO ICT Centre*)Sarvnaz Karimi (*CSIRO ICT Centre*)Bella Robinson (*CSIRO ICT Centre*)Mark Cameron (*CSIRO ICT Centre*)**Cager: A Framework for Cross-page Search** (Page 2704)Zhumin Chen (*Shandong University*)Byron J. Gao (*Texas State University*)Qi Kang (*Shandong University*)**Mixed-Initiative Conversational System Using Question-Answer Pairs Mined from the Web** (Page 2707)Wilson Wong (*RMIT University*)Lawrence Cavedon (*RMIT University*)John Thangarajah (*RMIT University*)Lin Padgham (*RMIT University*)**PicAlert!: A System for Privacy-Aware Image Classification and Retrieval** (Page 2710)Sergej Zerr (*University of Hannover*)Stefan Siersdorfer (*University of Hannover*)Jonathon Hare (*University of Southampton*)**TASE: A Time-aware Search Engine** (Page 2713)Sheng Lin (*University of Science and Technology of China*)Peiquan Jin (*University of Science and Technology of China*)Xujian Zhao (*University of Science and Technology of China*)Lihua Yue (*University of Science and Technology of China*)[\(Return to Top\)](#)**Gumshoe Quality Toolkit: Administering Programmable Search** (Page 2716)Zhuowei Bao (*University of Pennsylvania*)Benny Kimelfeld (*IBM Research - Almaden*)Yunyao Li (*IBM Research - Almaden*)Sriram Raghavan (*IBM India Research Lab*)Huahai Yang (*IBM Research - Almaden*)**Simultaneous Realization of Page-centric Communication and Search** (Page 2719)Yuhki Shiraishi (*Kyoto Sangyo University*)Jianwei Zhang (*Tsukuba University of Technology*)Yukiko Kawai (*Kyoto Sangyo University*)Toyokazu Akiyama (*Kyoto Sangyo University*)**MOUNA: Mining Opinions to Unveil Neglected Arguments** (Page 2722)Mouna Kacimi (*Free University of Bozen-Bolzano*)Johann Gamper (*Free University of Bozen-Bolzano*)**Databases Demonstration Session****MAGIK: Managing Completeness of Data** (Page 2725)Ognjen Savkovic (*Free University of Bozen-Bolzano*)Mirza Paramita (*Free University of Bozen-Bolzano*)Sergey Paramonov (*Free University of Bozen-Bolzano*)Werner Nutt (*Free University of Bozen-Bolzano*)**Exploration of Monte-Carlo based Probabilistic Query Processing in Uncertain Graphs** (Page 2728)Tobias Emrich (*Ludwig-Maximilians Universität*)Hans-Peter Kriegel (*Ludwig-Maximilians Universität*)Johannes Niedermayer (*Ludwig-Maximilians Universität*)Matthias Renz (*Ludwig-Maximilians Universität*)André Suhartha (*Ludwig-Maximilians Universität*)Andreas Züfle (*Ludwig-Maximilians Universität*)[\(Return to Top\)](#)**The Nautilus Analyzer: Understanding and Debugging Data Transformations** (Page 2731)Melanie Herschel (*Université Paris Sud / INRIA Saclay*)Hanno Eichelberger (*Universität Tübingen*)**Demonstrating ProApproX 2.0: A Predictive Query Engine for Probabilistic XML** (Page 2734)Asma Souihli (*Institut Mines-Télécom*)Pierre Senellart (*Institut Mines-Télécom*)**HadoopXML: A Suite for Parallel Processing of Massive XML Data with Multiple Twig Pattern Queries** (Page 2737)Hyeobong Choi (*Korea Advanced Institute of Science and Technology*)Kyong-Ha Lee (*Korea Advanced Institute of Science and Technology*)Soo-Hyong Kim (*Korea Advanced Institute of Science and Technology*)Yoon-Joon Lee (*Korea Advanced Institute of Science and Technology*)Bongki Moon (*University of Arizona*)**MADden: Query-Driven Statistical Text Analytics** (Page 2740)

Christan Grant (*University of Florida*)
 Joir-dan Gumbs (*University of Florida*)
 Kun Li (*University of Florida*)
 Daisy Zhe Wang (*University of Florida*)
 George Chitouras (*Greenplum/EMC*)

STFMap: Query- and Feature-Driven Visualization of Large Time Series Data Sets (Page 2743)

K. Selçuk Candan (*Arizona State University*)
 Rosaria Rossini (*University of Torino*)
 Maria Luisa Sapino (*University of Torino*)
 Xiaolan Wang (*Arizona State University*)

(Return to Top)

Primates: A Privacy Management System for Social Networks (Page 2746)

Inen Ben Dhia (*Institut Mines-Télécom, Télécom ParisTech*)
 Talel Abdesslem (*Institut Mines-Télécom, Télécom ParisTech*)
 Mauro Sozio (*Institut Mines-Télécom, Télécom ParisTech*)

AMANDA: Web Data Repositories in the Amazon Cloud (Page 2749)

Andrés Aranda-Andújar (*Inria Saclay*)
 Francesca Bugiotti (*Inria Saclay & Università Roma Tre*)
 Jesús Camacho-Rodríguez (*Inria Saclay & Université Paris-Sud*)
 Dario Colazzo (*Inria Saclay & Université Paris-Sud*)
 François Goasdoué (*Inria Saclay & Université Paris-Sud*)
 Zoi Kaoudi (*Inria Saclay & Université Paris-Sud*)
 Ioana Manolescu (*Inria Saclay & Université Paris-Sud*)

Workshop Summaries

DUBMMSM'12: International Workshop on Data-Driven User Behavioral Modeling and Mining from Social Media (Page 2752)

(Return to Top)

Jalal Mahmud (*IBM Research - Almaden*)
 Jeffrey Nichols (*IBM Research - Almaden*)
 James Caverlee (*Texas A&M University*)
 John O'Donovan (*University of California, Santa Barbara*)
 Michelle Zhou (*IBM Research - Almaden*)

CloudDB 2012: Fourth International Workshop on Cloud Data Management (Page 2754)

Xiaofeng Meng (*Renmin University of China*)
 Adam Silberstein (*LinkedIn*)
 Fusheng Wang (*Emory University*)

CDMW'12 - City Data Management Workshop: Workshop Summary (Page 2756)

Veli Bicer (*IBM Research - Ireland*)
 Thanh Tran (*Institute AIFB*)
 Fatma Ozcan (*IBM Research - Almaden*)
 Opher Etzion (*IBM Research - Haifa*)

Managing Interoperability and complexXity in Health Systems - MIXHS'12 (Page 2758)

Cui Tao (*Mayo Clinic*)
 Matt-Mouley Bouamrane (*The University of Glasgow*)

The 2012 International Workshop on Web-Scale Knowledge Representation, Retrieval, and Reasoning (Page 2760)

(Return to Top)

Spyros Kotoulas (*IBM Research*)
 Yi Zeng (*Chinese Academy of Sciences*)
 Zhisheng Huang (*VU University Amsterdam*)

SHB 2012: International Workshop on Smart Health and Wellbeing (Page 2762)

Christopher C. Yang (*Drexel University*)
 Hsinchun Chen (*University of Arizona*)
 Howard Wactlar (*National Science Foundation*)
 Carlo K. Combi (*Università degli Studi di Verona*)
 Xuning Tang (*Drexel University*)

BooksOnline'12: 5th Workshop on Online Books, Complementary Social Media and Their Impact (Page 2764)

Gabriella Kazai (*Microsoft Research*)
 Monica Landoni (*University of Lugano*)
 Carsten Eickhoff (*Delft University of Technology*)
 Peter Brusilovsky (*University of Pittsburgh*)

DTMBIO 2012: International Workshop on Data and Text Mining in Biomedical Informatics (Page 2766)

Sophia Ananiadou (*University of Manchester*)
 Doheon Lee (*Korea Advanced Institute of Science and Technology*)
 Hua Xu (*Vanderbilt University*)
 Min Song (*Yonsei University*)

PLEAD 2012: Politics, Elections and Data (Page 2768)

Ingmar Weber (*Yahoo! Research*)
 Ana-Maria Popescu (*Research Consulting*)
 Marco Pennacchiotti (*eBay Inc.*)

Workshop on Multimodal Crowd Sensing (CrowdSens 2012) (Page 2770)

Haggai Roitman (*IBM Research Haifa*)
Iván Cantador (*Universidad Autónoma de Madrid*)
Miriam Fernández (*Knowledge Media Institute*)

Fifth Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR'12): CIKM 2012 Workshop (Page 2772)

Jaap Kamps (*University of Amsterdam*)
Jussi Karlgren (*Gavagai*)
Peter Mika (*Yahoo! Research*)
Vanessa Murdock (*Microsoft Bing*)

[\(Return to Top\)](#)

First International Workshop on Information and Knowledge Management for Developing Regions (Page 2774)

Rakesh Agrawal (*Microsoft Search Labs*)
Douglas W. Oard (*University of Maryland*)
Nitendra Rajput (*IBM India Research Laboratory*)

PIKM 2012: 5th ACM Workshop for PhD Students in Information and Knowledge Management (Page 2776)

Aparna S. Varde (*Montclair State University*)
Fabian Suchanek (*Max Planck Institute for Informatics*)

WIDM'12: The 12th International Workshop on Web Information and Data Management (Page 2778)

George H. L. Fletcher (*Eindhoven University of Technology*)
Prasenjit Mitra (*The Pennsylvania State University*)

DOLAP 2012 Workshop Summary (Page 2780)

Matteo Golfarelli (*University of Bologna*)
Il-Yeol Song (*Drexel University*)