

# ObjectRank: Ranked Keyword Searching of Databases

Tech ID: 21169 / UC Case 2004-244-0

## BACKGROUND

The ObjectRank system applies the random-walk model—the effectiveness of which is proven by [Google's PageRank](#)—to keyword search in databases modeled as labeled graphs. The system ranks the database objects with respect to the user-provided keywords.

The PageRank technique assigns to each page ( $p$ ) a score that is based on the score of the pages pointing to  $p$ . Hence, pages pointed to by many high-score pages receive a high score as well. Alternatively, the score of  $p$  is equal to the probability that a random surfer, starting from a random page, will be at  $p$  at a specific time.

## TECHNOLOGY DESCRIPTION

We applied this idea to databases, which required modifications of the original algorithms to exploit the semantic information of every application, which is represented by annotating the database schema. Another novelty of ObjectRank is that we perform keyword-specific ObjectRank—not global, as Google does. Hence, for each <keyword, object> we compute an ObjectRank value. In particular, random walks start from the objects containing the keywords. Each object is ranked with respect to the particular keywords, based on the stationary probability that random walks are found at the object at a given time.

ObjectRank is adjustable to the semantics of each database and the varying requirements of object ranking. In particular, we adjust the weight of global importance, the weight of each keyword of the query, the importance of a result actually containing the keywords versus being referenced by nodes with high ObjectRank, and the volume of authority flow via each type of semantic connection.

Novel performance challenges and opportunities are addressed. First, schemas impose constraints on the graph, which are exploited for performance purposes. Second, we employ aggressive precomputation and experiment on storage space versus execution time tradeoffs. Finally, the keyword specific ObjectRanks are efficiently combined in order to produce the top- $k$  objects.

## APPLICATIONS

The following figure illustrates a small subset of the DBLP database in the form of a labeled graph (author, conference, and year nodes except for "R. Agrawal," "ICDE," and "ICDE 1997" respectively, are omitted to simplify the figure). The weights on edges represent the authority flow through them. For example, more authority is transferred from a paper to a cited paper than from a paper to its author. Also notice that no authority is transferred from the cited papers to the citing.

## CONTACT

University of California, San Diego  
Office of Innovation and Commercialization  
[innovation@ucsd.edu](mailto:innovation@ucsd.edu)  
tel: 858.534.5815.



## OTHER INFORMATION

### CATEGORIZED AS

- [Research Tools](#)
- [Other](#)

### RELATED CASES

2004-244-0

Given a keyword query, (e.g., the single keyword query "OLAP"), ObjectRank sorts the database objects by their importance with respect to the user-provided keywords. The papers of the above figure are ranked as follows:

Data Cube Modeling Multidimensional Databases Range Queries in OLAP Index Selection for OLAP

Notice that the "Data Cube" and the "Modeling Multidimensional Databases" papers do not contain the keyword "OLAP," but they clearly constitute important papers in the OLAP area, since they are referenced by other papers of the OLAP area, or have been written by authors who have written other important "OLAP" papers, or appear in conferences important to "OLAP."

RELATED MATERIALS

- ▶ Vagelis Hristidis, Yannis Papakonstantinou, Ramakrishna Varadarajan. [Using Proximity Search to Estimate Authority Flow](#). IEEE TKDE 2009.
- ▶ Ramakrishna Varadarajan, Vagelis Hristidis, Louiqa Raschid. [Explaining and Reformulating Authority Flow Queries](#). IEEE ICDE 2008.
- ▶ Vagelis Hristidis, Heasoo Hwang, Yannis Papakonstantinou. [Authority-Based Keyword Search in Databases](#). ACM Trans. Database Syst (TODS) 2008.
- ▶ Heasoo Hwang, Vagelis Hristidis, Yannis Papakonstantinou. ObjectRank: A System for Authority-Based Search on Databases. Demo paper, ACM SIGMOD 2006.
- ▶ Andrey Balmin, Vagelis Hristidis, Yannis Papakonstantinou: [Authority-Based Keyword Queries in Databases Using ObjectRank](#). VLDB 2004.
- ▶ [Random Walk in Stuctured Databases](#). UCSD Research Seminar, 2003.
- ▶ [Explaining and Reformulating Authority Flow Queries](#). IEEE ICDE, 2008.
- ▶ <http://db.ucsd.edu/ObjectRank/>

On the ObjectRank Demo Web page (see [http://db.ucsd.edu/BibObjectRank/main05\\_new.html](http://db.ucsd.edu/BibObjectRank/main05_new.html)), we describe some typical search profiles for various selections of parameters. A description of the demonstration is detailed below:

The DBLP subset of the demo consists of the available publications in 12 major database conferences, including SIGMOD, VLDB, PODS, ICDE, ICDT, and EDBT. The schema of the database is the following.

For AND (OR) semantics, the score of a node *u* is the product (sum) of the keyword-specific ObjectRanks of *u* with respect to each of the user-specified keywords. The demo also provides two other adjusting parameters: the "containment of actual keywords" parameter determines how important it is for a result to contain the user-specified keywords, as opposed to being pointed by other nodes with high score. The "global ObjectRank" parameter specifies if the global (keyword-independent) ObjectRank is included in the calculation of the scores.

PATENT STATUS

Country	Type	Number	Dated	Case
United States Of America	Issued Patent	8,862,594	10/14/2014	2004-244
United States Of America	Issued Patent	7,698,267	04/13/2010	2004-244

University of California, San Diego  
Office of Innovation and Commercialization  
9500 Gilman Drive, MC 0910, ,  
La Jolla, CA 92093-0910

Tel: 858.534.5815  
innovation@ucsd.edu  
<https://innovation.ucsd.edu>  
Fax: 858.534.7345

© 2010 - 2014, The  
Regents of the University of  
California  
[Terms of use](#)  
[Privacy Notice](#)