



## The thematic and citation landscape of *Data and Knowledge Engineering* (1985–2007) <sup>☆</sup>

Chaomei Chen <sup>a</sup>, Il-Yeol Song <sup>a,\*</sup>, Xiaojun Yuan <sup>b</sup>, Jian Zhang <sup>a</sup>

<sup>a</sup> College of Information Science and Technology, Drexel University, United States

<sup>b</sup> College of Computing and Information, University at Albany, State University of New York, United States

### ARTICLE INFO

#### Article history:

Available online 28 May 2008

#### Keywords:

Structural and temporal patterns  
Domain analysis  
Scientometrics  
CiteSpace  
Thematic analysis  
DKE

### ABSTRACT

The thematic and citation structures of *Data and Knowledge Engineering* (DKE) (1985–2007) are identified based on text analysis and citation analysis of the bibliographic records of full papers published in the journal. Temporal patterns are identified by detecting abrupt increases of frequencies of noun phrases extracted from titles and abstracts of DKE papers over time. Conceptual structures of the subject domain are identified by clustering analysis. Concept maps and network visualizations are presented to illustrate salient patterns and emerging thematic trends. A variety of statistics are reported to highlight key contributors and DKE papers that have made profound impacts.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

Data and Knowledge Engineering (DKE) is a premier journal focusing on the common research themes in the areas of database systems and knowledge base systems. The major aim of the journal is to identify, investigate and analyze the underlying principles in the design and effective use of these systems. DKE publishes original research results, technical advances, and new experiments concerning data engineering, knowledge engineering, and the interface of these two fields. Under the leadership of Professors Peter Chen and Reind van de Riet, DKE has served its roles as a leading scholarly journal.

Since its first publication in 1985, DKE has reached a world-wide audience of researchers, designers, managers and users. DKE also has covered a wide spectrum of research topics such as representation and manipulation of data and knowledge, architectures of database and knowledge-based systems, construction of data and knowledge bases, tools and methodologies for developing data and knowledge bases, applications, case studies, and management issues. DKE has made significant contributions and impacts on the advances of data and knowledge engineering.

In order to celebrate the 25th anniversary of DKE, we embarked an extensive analysis of papers that have been published in DKE over the last 24 years. Our analysis includes authorship analysis, coauthorship analysis, thematic trends and conceptual structures, visualization of topic maps and various citation networks. The analysis is based on bibliographic records and citation information retrieved from the Web of Science (WoS), Scopus (<http://www.scopus.com/>), and ScienceDirect (<http://www.sciencedirect.com/>).

<sup>☆</sup> CiteSpace is available at <http://cluster.cis.drexel.edu/~cchen/citespace>. Color versions of the figures are available at <http://cluster.cis.drexel.edu/~cchen/papers/2008/dke/figures>.

\* Corresponding author.

E-mail addresses: [chaomei.chen@cis.drexel.edu](mailto:chaomei.chen@cis.drexel.edu) (C. Chen), [songiy@drexel.edu](mailto:songiy@drexel.edu) (I.-Y. Song), [xy454777@albany.edu](mailto:xy454777@albany.edu) (X. Yuan), [jz85@drexel.edu](mailto:jz85@drexel.edu) (J. Zhang).

## 2. A longitudinal study of DKE

Our study primarily utilizes CiteSpace<sup>1</sup> as an analytic and visualization tool. CiteSpace is a freely available Java application for analyzing and visualizing emerging trends and citation patterns in scientific literature [1,2]. CiteSpace is continuously evolving to incorporate a variety of visual analytic functions. CiteSpace is designed to simplify the analysis of scientific literature by enabling users to find salient patterns from a diverse range of visual attributes. CiteSpace follows a simple model of the dynamics of scholarly communication in which a transient body of scientific papers, collectively known as a research front, makes reference to a group of papers in the literature, which is called the intellectual basis [1]. Based on this model, CiteSpace aims to make it easy for users to identify some special classes of papers in terms of *landmarks* by citation popularity, *hotspots* by abrupt increases of citations they received, and *pivotal papers* that are strategically positioned in co-citation networks. Landmark papers are depicted by their large-sized tree-ring circles. Hotspots are shown as nodes with a red rim. Pivotal papers are shown as nodes with a purple rim.

The general procedure of analysis and visualization with CiteSpace is outlined as follows. Technical details are provided elsewhere [1,2].

1. Identify a knowledge domain. In this study, the knowledge domain is defined by DKE papers and their citations.
2. Data collection. We collect bibliographic records and citation data associated with DKE papers from three sources: ScienceDirect,<sup>2</sup> Scopus,<sup>3</sup> and the Web of Science.<sup>4</sup> See Section 3.1 for detail.
3. Extract noun phrase terms from titles, abstracts, descriptors, and identifiers of citing articles in the dataset. The burstness of the extracted terms is detected for abruptly increased frequencies of specific terms [3]. Burst terms are used to capture fast-growing interests.
4. Time slicing. Specify the range of the entire time interval and the length of a single time slice.
5. Threshold selection. CiteSpace allows users to specify three sets of threshold levels for citation counts, co-citation counts, and co-citation coefficients. Citation counts are the number of times a publication is cited by DKE papers. Two publications are called co-cited if a paper cites both of them. Co-citation counts for a given pair of publications are the number of papers in our dataset that cite the pair. Co-citation coefficients are normalized co-citation counts over each time slice. The specified thresholds are applied to three time slices, namely, the earliest slice, the middle one, and the last one. Linear interpolated thresholds are assigned to the rest of slices. CiteSpace supports networks of four types of nodes and three types of links. Nodes include authors, papers, journals, and burst terms, whereas links may represent co-occurrence, co-citation, or referential links.
6. Pruning and merging. Pathfinder network scaling [4,5] is the default option in CiteSpace for network pruning [2,4]. Users choose whether or not to apply the scaling operation to individual networks. CiteSpace merges individual networks by taking a set union of all the vertices and selecting links that do not violate a triangle inequality condition in overlapping areas between networks. Users can choose whether or not to prune the merged network as a whole.
7. Layout. CiteSpace supports a standard graph view and a time-zone view.
8. Visual inspection. CiteSpace enables users to interact with the visualization of a knowledge domain in several ways. The user may control the display of visual attributes and labels as well as a variety of parameters used by the underlying layout algorithms.
9. Verify pivotal points. The significance of a marked pivotal point can be verified by asking domain experts, for example, the authors of pivotal-point articles, and/or examining the literature, such as passages containing citations of a pivotal-point article. A particularly interesting direction of research is the development of tools that can automatically summarize the value of a pivotal point. Digital libraries, automated text summarization, machine learning, and several other fields are among the most promising sources of input.

## 3. Methods

### 3.1. Data collection

We searched for bibliographic records of the journal *Data & Knowledge Engineering* (DKE) in three sources, namely ScienceDirect, Scopus, and the Web of Science. ScienceDirect has the most extensive coverage of papers published in DKE. We found 993 bibliographic records of DKE papers in ScienceDirect, 865 DKE records in Scopus, and 774 records in the Web of Science.

The three sources vary in their accessibility of cited references. Although users can access cited references in interactive modes, the Web of Science provides the most reliable download function. The Web of Science imposes a 500-record limit per download request. Multiple download operations are necessary to retrieve a dataset with more than 500 records. Scopus

<sup>1</sup> <http://cluster.cis.drexel.edu/~cchen/citespace/>.

<sup>2</sup> <http://www.sciencedirect.com/>.

<sup>3</sup> <http://www.scopus.com/>.

<sup>4</sup> <http://scientific.thomson.com/products/wos/>.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات