# Full-fledged semantic indexing and querying model designed for seamless integration in legacy RDBMS(Article)

- Tekli, J.[a]Email Author,
- Chbeir, R.[b],
- Traina, A.J.M.[c],
- Traina, C., Jr.[c],
- Yetongnon, K.[d],
- Ibanez, C.R.[e],
- Al Assad, M.[a],
- Kallas, C.[a]
- View Correspondence (jump link)

- [a]Lebanese American University, ECE Dept., Byblos, Lebanon
- [b]University of Pau & Pays Adour, LIUPPA Lab., Anglet, France
- [c]University of Sao Paulo, ICMC, Sao Carlos, Brazil

View additional affiliations

## Abstract View references (101)

In the past decade, there has been an increasing need for semantic-aware data search and indexing in textual (structured and NoSQL) databases, as full-text search systems became available to non-experts where users have no knowledge about the data being searched and often formulate query keywords which are different from those used by the authors in indexing relevant documents, thus producing noisy and sometimes irrelevant results. In this paper, we address the problem of semantic-aware querying and provide a general framework for modeling and processing semantic-based keyword queries in textual databases, i.e., considering the lexical and semantic similarities/disparities when matching user query and data index terms. To do so, we design and construct a semantic-aware inverted index structure called SemIndex, extending the standard inverted index by constructing a tightly coupled inverted index graph that combines two main resources: a semantic network and a standard inverted index on a collection of textual data. We then provide a general keyword query model with specially tailored query processing algorithms built on top of SemIndex, in order to produce semantic-aware results, allowing the user to choose the results' semantic coverage and expressiveness based on her needs. To investigate the practicality and effectiveness of SemIndex, we discuss its physical design within a standard commercial RDBMS allowing to create, store, and query its graph structure, thus enabling the system to easily scale up and handle large volumes of data. We have conducted a battery of experiments to test the performance of SemIndex, evaluating its construction time, storage size, query processing time, and result quality, in comparison with legacy inverted index. Results highlight both the effectiveness and scalability of our approach. © 2018

## Author keywords

Inverted indexNoSQL indexingSemantic networkSemantic queriesSemantic-aware data processingTextual databases

## Indexed keywords

| | |
|---|---|
| **Engineering controlled terms:** | Data handlingDigital storageIndexing (materials working)Indexing (of information)Query languagesQuery processingSemantics |
| **Engineering uncontrolled terms** | Inverted indicesSemantic networkSemantic querySemantic-awareTextual database |
| **Engineering main heading:** | Search engines |