

Search-As-You-Type in Forms:

Leveraging the Usability and the Functionality
of Search Paradigm in Relational Databases



Hao Wu

Supervised by Prof. Lizhu Zhou

Database Research Group, Tsinghua University

VLDB PhD Workshop - Sept. 13, Singapore

Motivation

Problem Statement

Challenges

Initial Achievements

Conclusions

Motivation

Problem Statement

Challenges

Initial Achievements

Conclusions

Motivation



- Relational databases are widely used.
- There are many search paradigms:
 - Structured Query Language (SQL)
 - Keyword Search (KS)
 - Query-By-Example (QBE)
- Different search paradigms are needed by different users.

Motivation

#1: SQL is complex.

```
SELECT *  
FROM Author A, Autor_Paper AP, Paper P  
WHERE title LIKE 'keyword' AND  
title LIKE 'search' AND  
authors LIKE 'g%' AND  
A.id = AP.aid AND  
P.id = AP.pid
```

Motivation

#2: Traditional keyword search is imprecise.

keyword search g



Title? Conf. name? Author name?

Motivation

#3: Form is awkward.

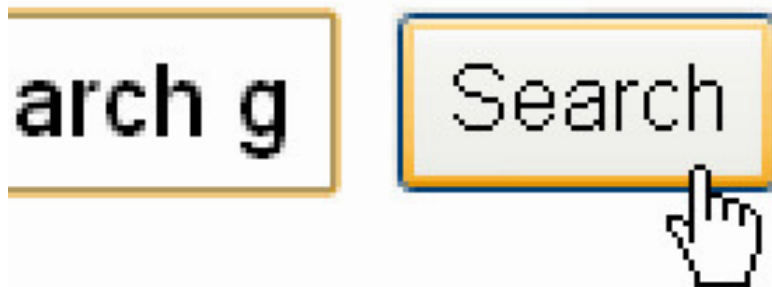
Search for People & Departments [[Basic Search](#)]

Search Group	Faculty, Staff, Students & Departments ▾	
First Name/Nickname	<input type="text"/>	Starts With ▾
Last Name	<input type="text"/>	Starts With ▾
E-mail	<input type="text"/>	Starts With ▾
UCInetID	<input type="text"/>	Starts With ▾
Department	<input type="text"/>	Starts With ▾
Telephone	<input type="text"/>	

UCI Directory: http://directory.uci.edu/index.php?form_type=advanced_search

Motivation

#4: The "Search" button is not convenient.



Motivation



Keyword Search
+ Form-Style Interface
+ Search-as-you-type
= **Seaform**



Motivation

Problem Statement

Challenges

Initial Achievements

Conclusions



Motivation

Problem Statement

Challenges

Initial Achievements

Conclusions

Problem Statement

- Data:
 - Single relational table.
 - Several searchable attributes.

ID	Title	Conf.	Author
1	xml database	VLDB	albert
2	xml database	SIGMOD	bob
3	xml search	VLDB	albert
4	xml security	VLDB	alice
5	rdbms	SIGMOD	charlie

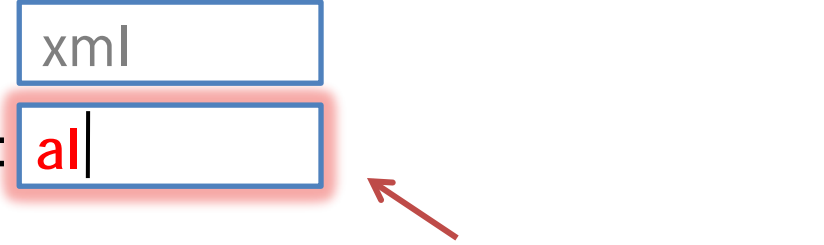
Problem Statement

- Query:
 - A set of keywords (prefixes) split by fields.
 - A focus indicator.

Title:

Author:

Focus = **Author**



Problem Statement

- Results:
 - **Global results:** corresponding tuples.
 - **Local results:** corresponding attribute values.
 - **Aggregations.**

Title:

Author:

albert	2
alice	1

xml database (albert)
xml search (albert)
xml security (alice)

✓ Motivation

✓ Problem Statement

Challenges

Initial Achievements

Conclusions

✓ Motivation

✓ Problem Statement

Challenges

Initial Achievements

Conclusions

Challenges: Search-As-You-Type

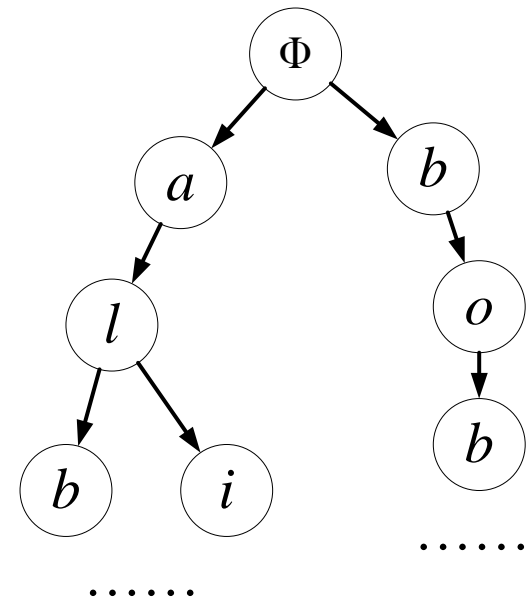
- Prefix matching:

- E.g. **al** → **al**bert, **al**ice, ...
Trie structure w/ cache.

- Fast response:

- Synchronization of local results and global results yields heavy computational cost.

On-demand synchronization and dual-list trie.



Challenges: Error Tolerance



- Misplacing of keywords:
 - E.g. input "albert" into the **Title** input box.
Automatic query refinement (given a query, how can we modify it to obtain more results?)
Large search space; rely on precise estimation and probabilistic model.
- Fuzzy matching:
 - E.g. input "albrt" instead of "albert".
Edit-distance computation on trie structure.
Ranking issue of local results: should local results be sorted by edit-distance, or by aggregation values?

Challenges: Scalability

- Handle large-scale databases:
 - There are large number of tuples.
 - 1) Top-k algorithm
Precise aggregation is impossible in this case.
 - 2) Using RDBMS itself
Index structure should be redesigned for DBMS; performance issues.
- Handle multiple tables:
 - Data are regularized to several tables.
Generalize the single-table local-global computation and reduce on-the-fly joins using pre-joined tables.
It is hard to determine which tables are the most necessary to pre-join; extra storage cost.

✓ Motivation

✓ Problem Statement

✓ Challenges

Initial Achievements

Conclusions

- ✓ Motivation
- ✓ Problem Statement
- ✓ Challenges

Initial Achievements

Conclusions

Initial Achievements

Seaform-DBLP

Features:

- Single table.
- Prefix matching.
- Average response time is less than 30 ms.

Limitations:

- Does not tolerate errors.
- Non-top-k, i.e. it returns all matching results.
- Memory-resident.

The screenshot shows the Seaform DBLP (beta) web interface. The search results are displayed in a table with the following data:

Author	Count
Guoliang Li	14
Rémi Gilleron	3
Ashish Goel	3
Lin Guo	3
Luis Gravano	3
Ning Gao	2
Jihong Guan	2
Konstantin Golenberg	2
Georgia Koutrika	2
Hector Garcia-Molina	2
Basilios Gatos	2

The interface also displays a list of search results with titles and authors, such as:

- [Finding and ranking compact connected trees for effective keyword proximity search in XML documents.](#) by Jianhua Feng; Guoliang Li; Jianyong Wang; Lizhu Zhou. *Inf. Syst.* 35(2): 186-203 (2010).
- [Searching RDF Graphs with SPARQL and Keywords.](#) by Shady Elbassuoni; Maya Ramanath; Ralf Schenkel; Gerhard Weikum. *IEEE Data Eng. Bull.* 33(1): 16-24 (2010).
- [Enhancing Keyword Search in Relational Databases Using Nearly Duplicate Records.](#) by Xiaochun Yang; Bin Wang; Guoren Wang; Ge Yu. *IEEE Data Eng. Bull.* 33(1): 60-66 (2010).
- [k Keyword Search in Relational Databases.](#) by Yanwei Xu; Yoshiharu Ishikawa; Jihong Guan. *WAIM 2010: 755-767*.
- [Automatically incorporating new sources in keyword search-based data integration.](#) by Partha Pratim Talukdar; Zachary G. Ives; Fernando

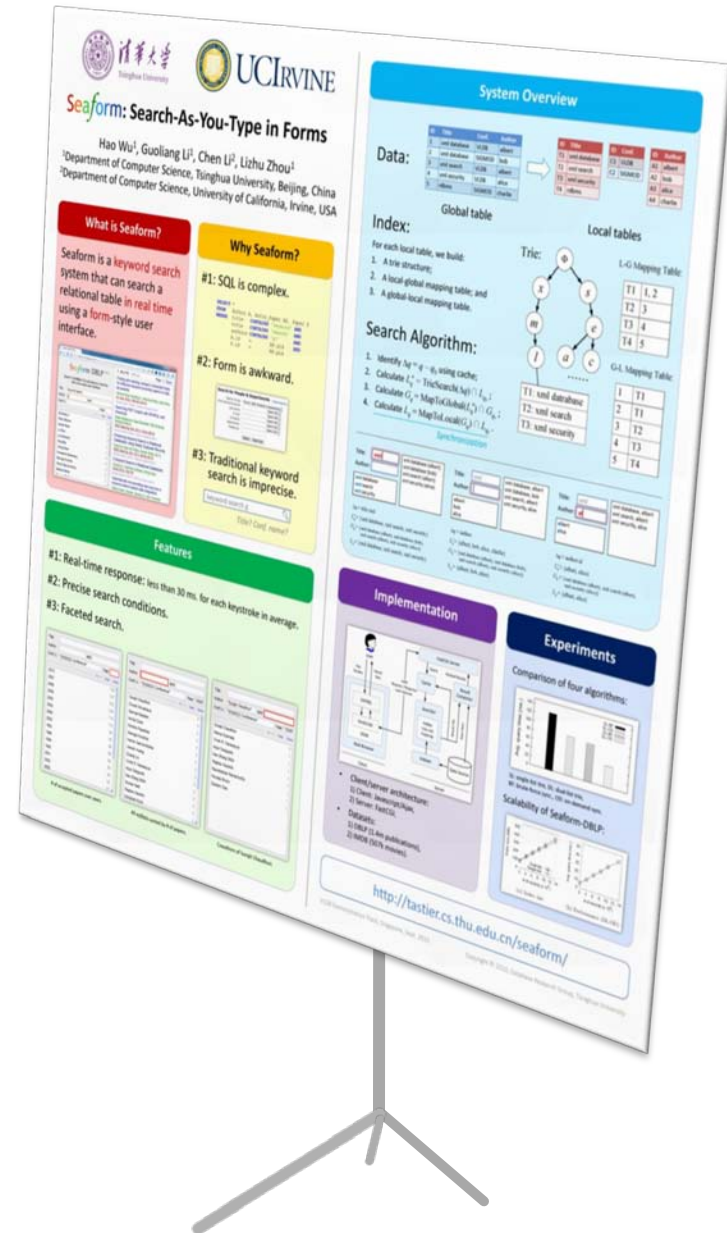
Demonstrations:

2

Sept. 14, Tuesday
14:00 to 15:30

5

Sept. 15, Wednesday
14:00 to 15:30



- ✓ Motivation
 - ✓ Problem Statement
 - ✓ Challenges
 - ✓ Initial Achievements
- Conclusions

- ✓ Motivation
- ✓ Problem Statement
- ✓ Challenges
- ✓ Initial Achievements

Conclusions

Conclusions

- Search-as-you-type with form is a good choice to balance the usability and functionality.
- There are still many problems to solve:
 - More effective index other than *trie + inverted lists*.
 - Support error tolerance.
 - Native DBMS support.
 - Top-k algorithms.
 - Pre-join (materialize) tables.
 - ...

Thanks

<http://tastier.cs.thu.edu.cn/seaform/>

My homepage: <http://dbgroup.cs.thu.edu.cn/wuhao/>