iQbees: Towards Interactive Semantic Entity Search Based on Maximal Aspects

Grzegorz Sobczak[1] Mateusz Chochół[2] Ralf Schenkel[3]
Marcin Sydow[1,2]

[1]Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

[2]Polish-Japanese Academy of Information Technology, Warsaw, Poland

[3]Universität Passau, Germany

22.10.2015

Table of contents

Similar Entity Search

Definition

Given a set of *query entities* Q, retrieve a ranked list of the k most similar entities $\mathcal{R}$.

Example

Let $Q = \{Saudi\ Arabia, Iraq\}$ and $k = 3$. A system returns other countries with large oil reserves:
$\mathcal{R} = \{Kuwait,\ Qatar, United\ Arab\ Emirates\}$.

Motivations

- QBEES (Query By Entity Example Search)
  - given set of entities find an entity that maximally resembles all of them (e.g. replacement of a particular person or part, etc.)
- IQBEES (Interactive QBEES)
  - the user iteratively selects (relevance feedback) example entities one by one to refine some concept represented by the entites

Knowledge Graph

Definition

A Knowledge Graph *KG* is a directed multi-graph that consists of three basic components, a *Fact Graph FG*, an *Ontology Tree O*, and a set of type assignment arcs *TA* connecting the two.

Notes

Arcs in *KG* are labelled. We will use the notation `relation`(`arg1`, `arg2`) for any directed arc with label `relation` in *KG* that points from node `arg1` to `arg2`.

Fact Graph

Definition

The fact graph $FG = (E, F)$ is a directed multigraph where nodes in E represent *entities* (e.g. `Chopin`, `Poland`) and edges in F represent *facts* about the entities.

Example

An arc `wasBornIn`(`Chopin`,`Poland`) represents the fact "Chopin was born in Poland".

Notes

The fact graph is a multi-graph, since there are possibly multiple parallel arcs between the same pair of entities (e.g. "Warsaw is the capital of Poland" and "Warsaw is the largest city in Poland").

Ontology Tree

Definition

The Ontology Tree $O = (C, S)$ is a graph where each node (class) $c \in C$ represents some *type* of entities (e.g. person). The class nodes are connected by directed arcs labelled as `subClassOf`.

Example

Triple `subClassOf(composer,musician)` indicates that every composer is also a musician.

Type Assignment

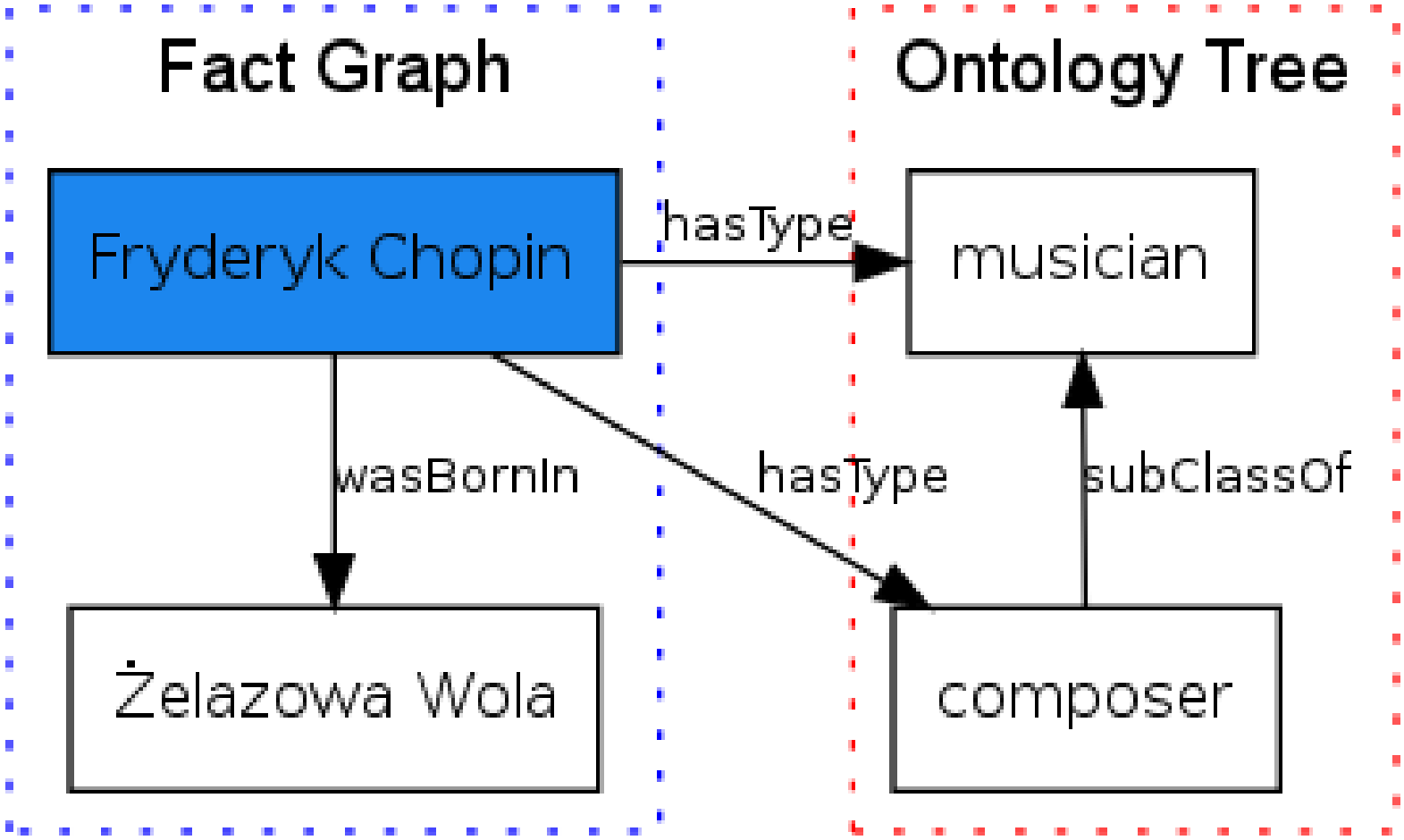
Definition

The type assignment *TA* is a set of arcs labelled `hasType` which connect entities from the Fact Graph end classes from the Ontology Tree. Each arc of the form `hasType(anEntity,aClass)` in *KG* means that the entity `anEntity` is an instance of the class `aClass`.

Example

For example the arc `hasType(Chopin,composer)` means that "Chopin is a composer".

Chopin example



Basic aspects (1/2)

Intuition

For any entity q, a basic aspect represents some "atomic property" of this entity (e.g. birthplace, type, occupation); the entity is characterized by the set of all "atomic properties".

Example

An entity `Chopin` (a famous Polish composer), is represented by the following "basic properties": "being born in Poland" and "being a composer".

Basic aspects (2/2)

Generalisation

By replacing the particular entity q in such an arc with a variable we obtain a *logical predicate* with one free variable.

Example

A factual arc `bornIn(Chopin,Poland)` and a type arc `hasType (Chopin,composer)` naturally induce predicates of the form `bornIn (.,Poland)` and `hasType(.,composer)` that represent the "basic properties" of this entity of "being born in Poland" and "being a composer", respectively.

Compound aspects

Definition

A set of basic aspects is called a *compound aspect*.

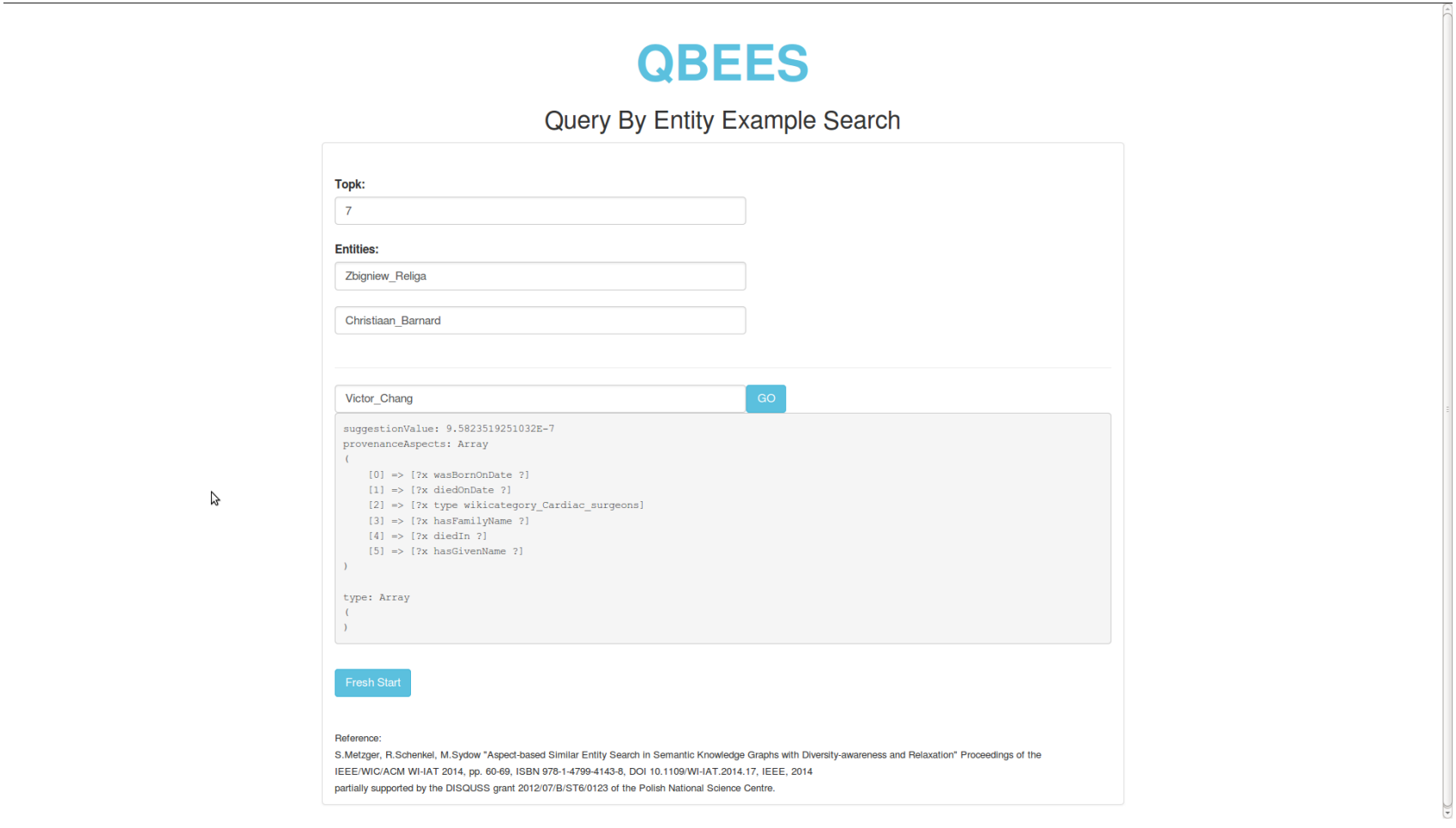
Example

A property "being a composer born in Poland", which consists of two "atomic properties" - "being a composer" and "being born in Poland", is represented by a compound aspect $\{\texttt{bornIn}(.,\texttt{Poland}), \texttt{hasType}(.,\texttt{composer})\}$.

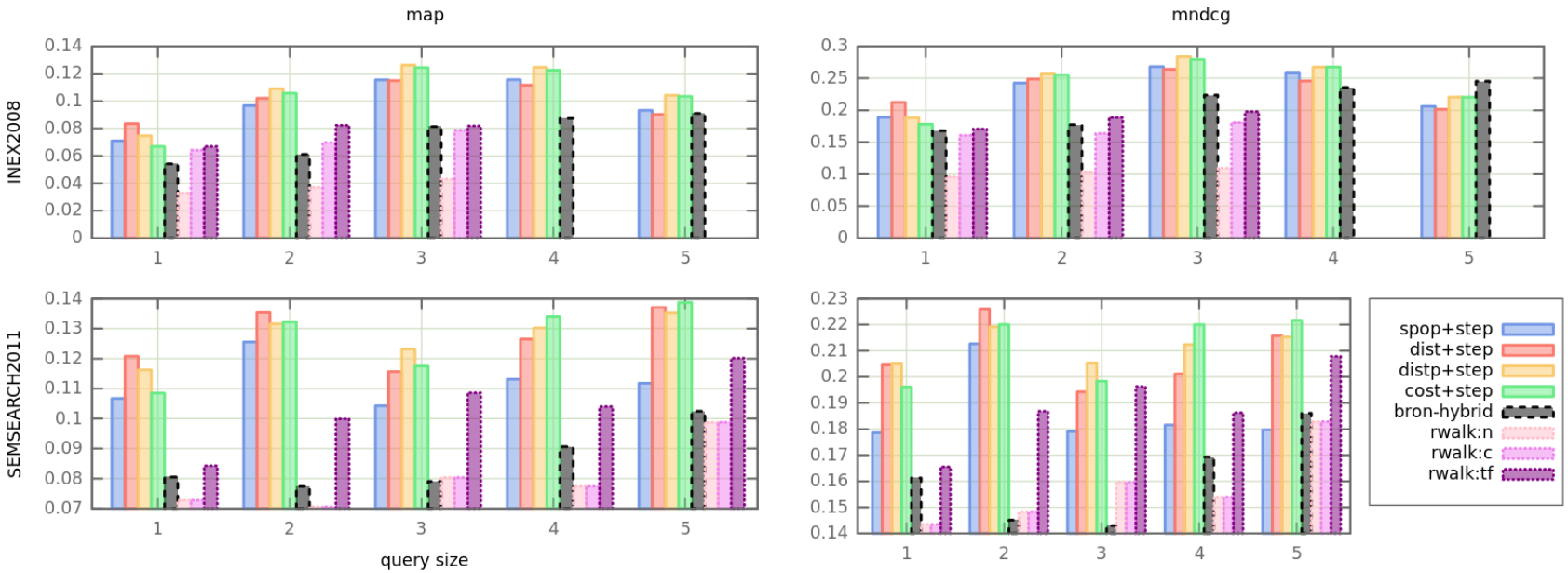
Maximal aspects

- Each entity can be treated as a set of basic aspects Let A_e be a set of all basic aspects of entity $e \in E$.
- Let q be an query example and A_q be its set of basic aspects.
- For all $e \in E, e \neq q$ consider set of all basic aspects common with q, that is $A'_e = A_e \cap A_q$.
- These compound aspects naturally form a lattice (with inclusion as an operation).
- *Maximal aspects* are those compound aspects which are maximal in the lattice.
- Entities that satisfy *maximal aspects* are returned as the most similar entities.

QBEES interface



QBEES evaluation



IQBEES

The procedure is as follows:

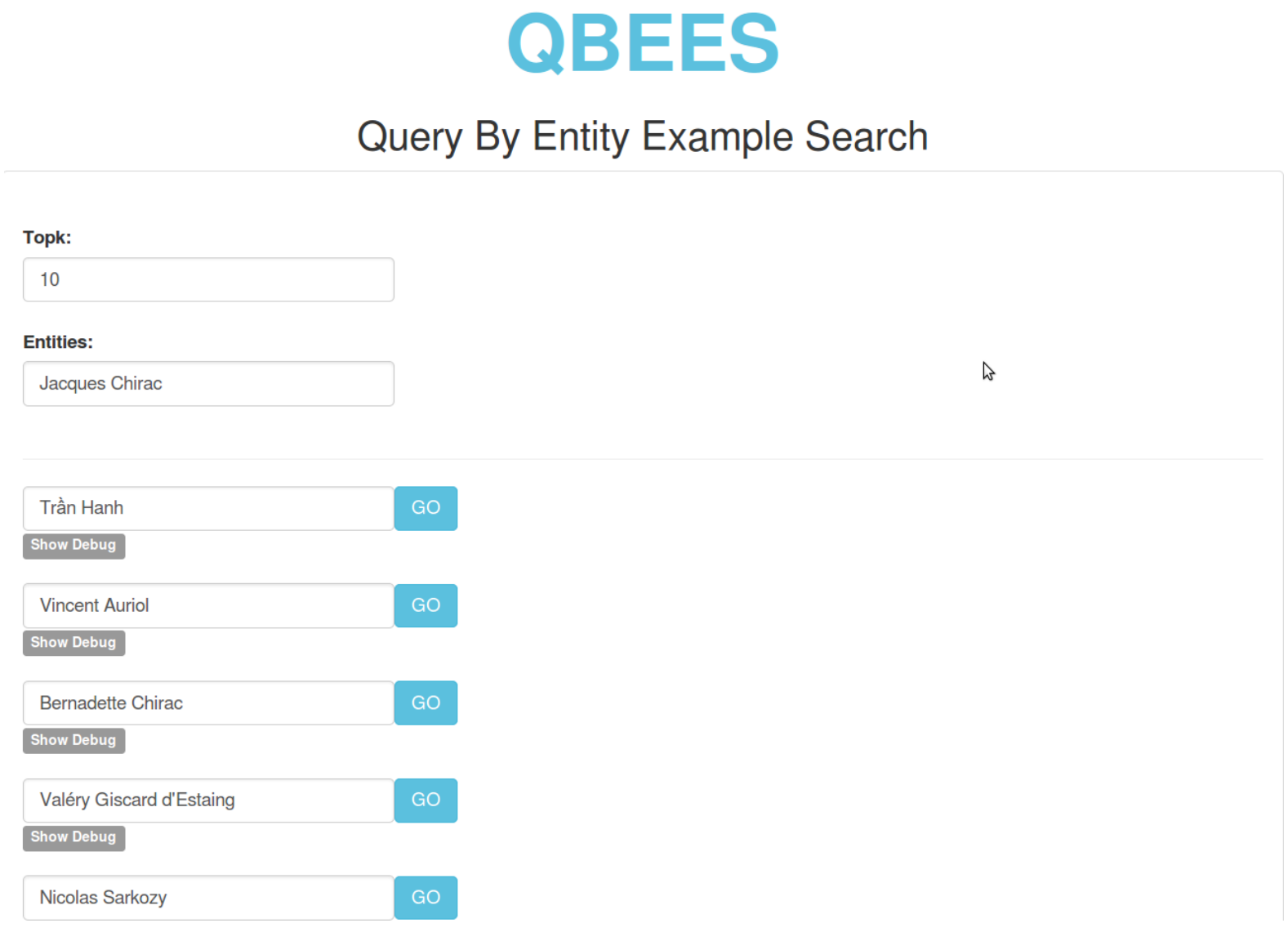
1. A user provides an initial example entity as the input.
2. The system returns a list of similar entities based on the QBEES approach.
3. If the results do not satisfy user information need, the user can treat the returned entities as *refinement suggestions* and select one of them as a hint for the system to refine his query. This entity is appended to the list of previously selected query entities. The user can go back to the point 2 until she finds the result successful or wishes to restart the search.

See the prototype under the following URL:
`http://webmining.pjwstk.edu.pl/iqbees_gui/`

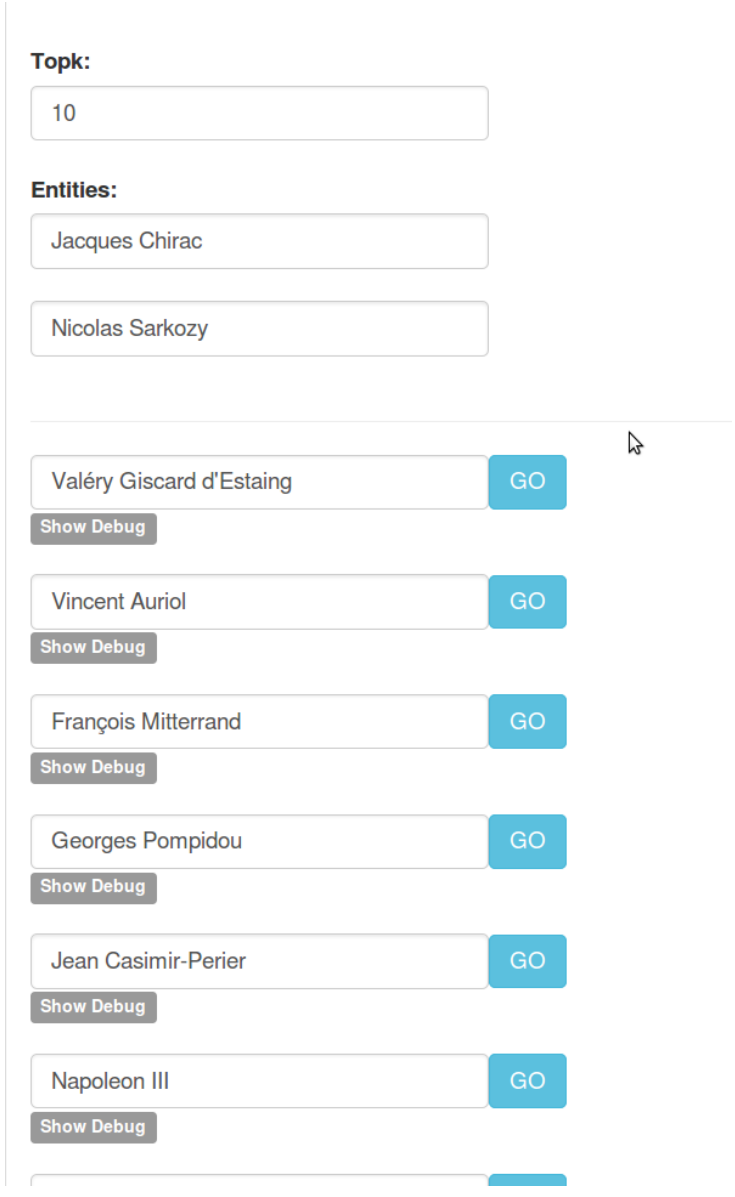
Examples (live demo)

- Jacques Chirac (Presidents of France)
- Paris (capitals of European countries)
- Vistula (rivers in Poland)

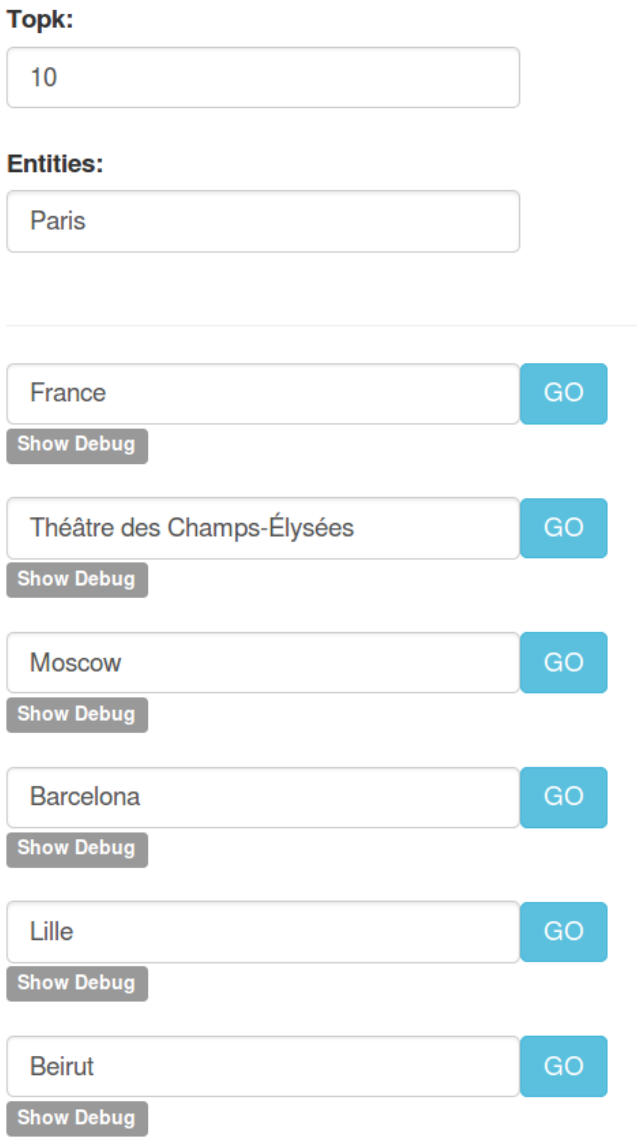
Jacques Chirac (1/2)



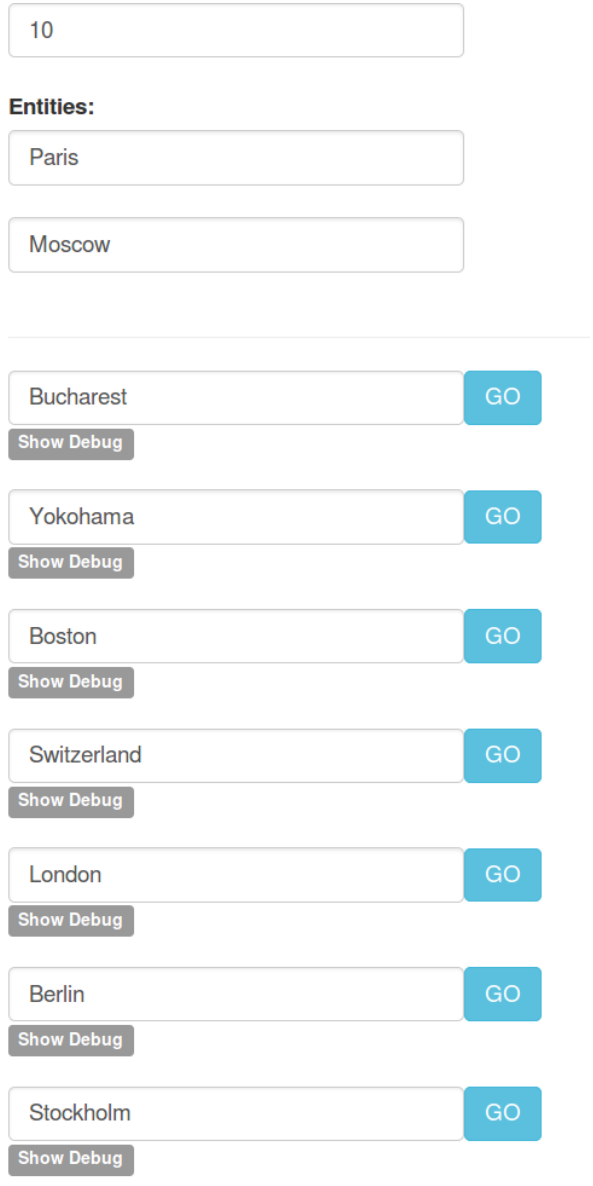
Jacques Chirac (2/2)



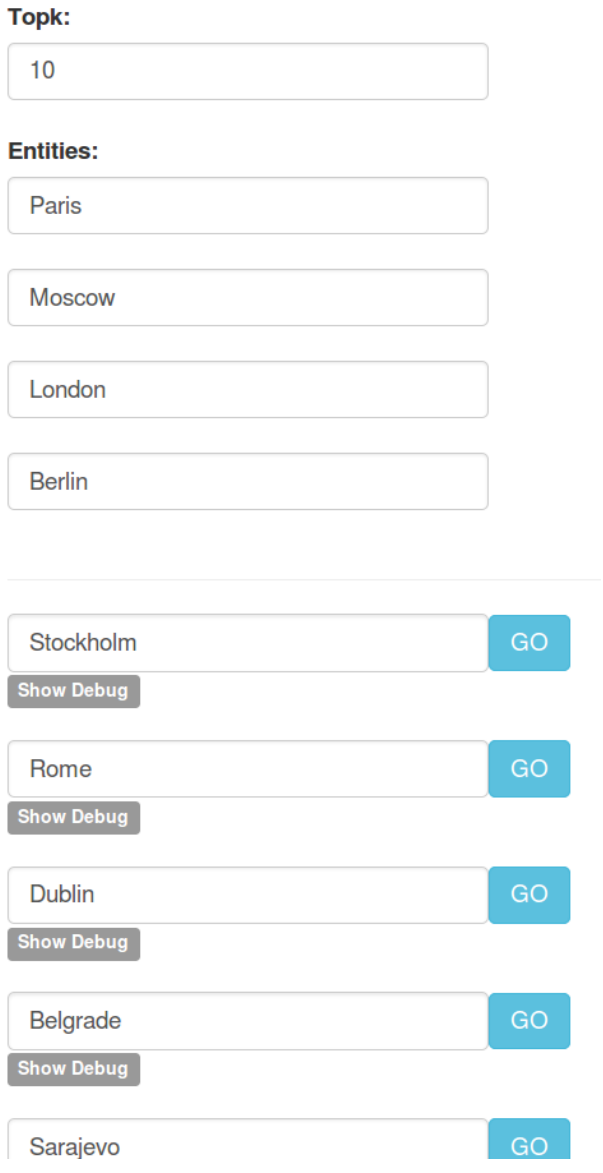
Paris (1/3)


Topk:
10
Entities:
Paris
France
GO
Show Debug
Théâtre des Champs-Élysées
GO
Show Debug
Moscow
GO
Show Debug
Barcelona
GO
Show Debug
Lille
GO
Show Debug
Beirut
GO
Show Debug

Paris (2/3)


10
Entities:
Paris
Moscow
Bucharest
GO
Show Debug
Yokohama
GO
Show Debug
Boston
GO
Show Debug
Switzerland
GO
Show Debug
London
GO
Show Debug
Berlin
GO
Show Debug
Stockholm
GO
Show Debug

Paris (3/3)


Topk:
10
Entities:
Paris
Moscow
London
Berlin
Stockholm
GO
Show Debug
Rome
GO
Show Debug
Dublin
GO
Show Debug
Belgrade
GO
Show Debug
Sarajevo
GO

Vistula


Topk:
10
Entities:
Vistula
Lusatian Neisse
GO
Show Debug
Warta
GO
Show Debug
Wel
GO
Show Debug
Bóbr
GO
Show Debug
Ner
GO
Show Debug
Rega
GO
Show Debug
Biebrza
GO
Show Debug

IBEES, Paris

Entities:

Paris

Negative Entities:

Suggestions:

France - +
Show Debug

Baghdad - +
Show Debug

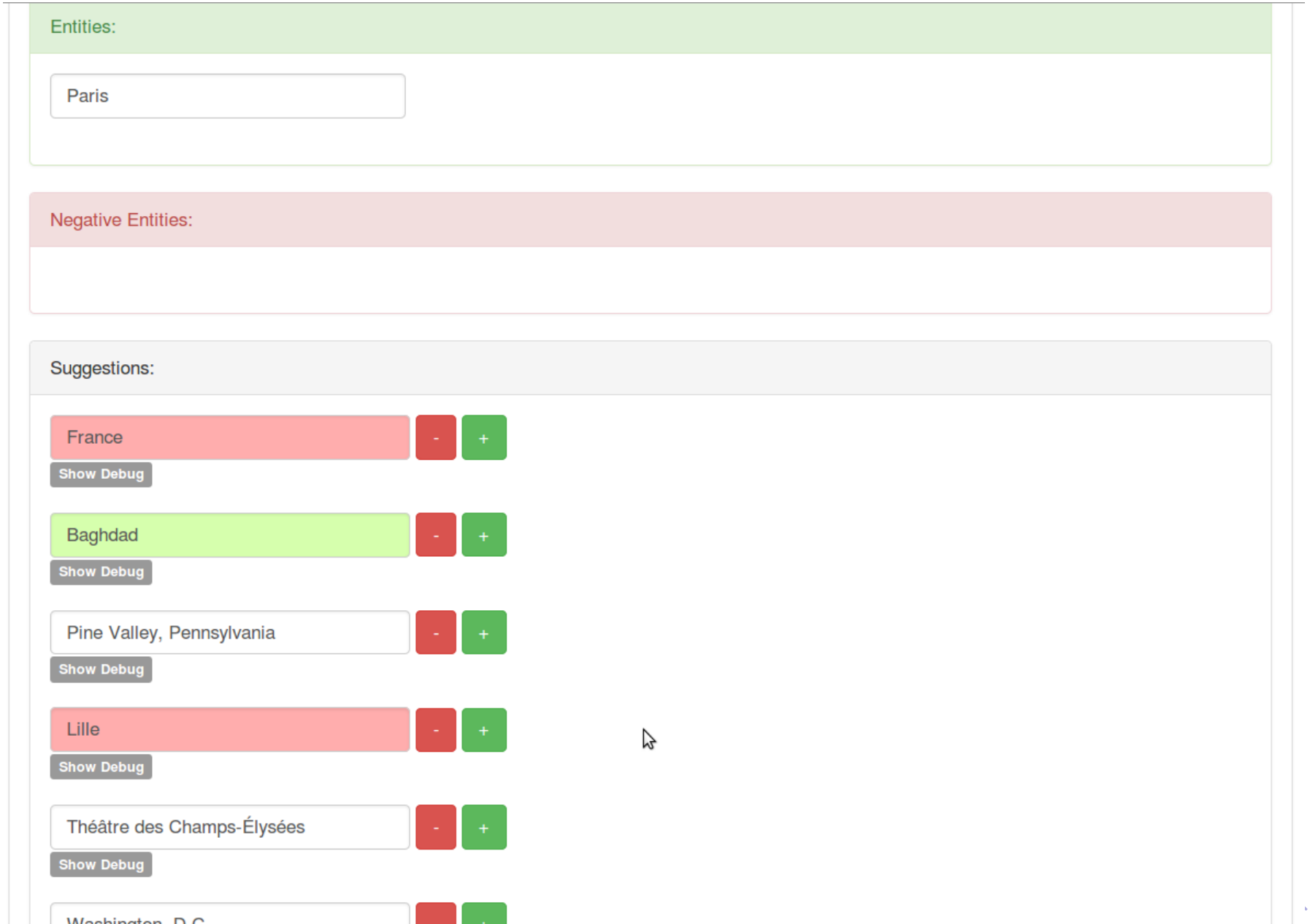
Pine Valley, Pennsylvania - +
Show Debug

Lille - +
Show Debug

Théâtre des Champs-Élysées - +
Show Debug

Washington, D.C. - +

IBEES, Paris


Entities:
Paris
Negative Entities:
Suggestions:
France
-
+
Show Debug
Baghdad
-
+
Show Debug
Pine Valley, Pennsylvania
-
+
Show Debug
Lille
-
+
Show Debug
Théâtre des Champs-Élysées
-
+
Show Debug

IBEES, Paris

Paris

Baghdad

Moscow

Negative Entities:

France

Lille

Suggestions:

London - +

Show Debug

Amsterdam - +

Show Debug

Berlin - +

Show Debug

Stockholm - +

Show Debug

Helsinki - +

Acknowledgements

This work is partially supported by:

- the Polish National Science Centre grant 2012/07/B/ST6/01239
- European Union under the European Social Fund Project PO KL "Information technologies: Research and their interdisciplinary applications", Agreement UDA-POKL.04.01.01-00-051/10-00.

Thank you