SQuID: Semantic Similarity-Aware Query Intent Discovery

Anna Fariha
afariha@cs.umass.edu
University of Massachusetts Amherst

Sheikh Muhammad Sarwar
smsarwar@cs.umass.edu
University of Massachusetts Amherst

Alexandra Meliou
ameli@cs.umass.edu
University of Massachusetts Amherst

Traditional data retrieval is challenging

- Non expert users struggle to formulate complex SQL query
- Traditional data retrieval is challenging for complex schema
- no SQL expertise required
- Example tuples: Iron Man, Spider Man, The Incredible Hulk

Query by Example (QBE) tries but fails

- But I wanted Marvel superhero movies only!
- Existing QBE approaches fail to capture semantic context (e.g., Marvel movies)

SQuID: a query intent discovery framework

- SQuID is aware of semantic similarity

Semantic context can be complex

- How to express funny?
- Implicit context
- Humans can tell easily!
- SQuID captures implicit semantic context

Semantic context: funny actors

- No explicit attribute, but hidden in the data
- Appearing in more than 40 comedy movies

SQuID in a nutshell

- Real-time performance through precomputing basic and derived semantic properties

Example Tuples

Iron Man
Spider Man
The Incredible Hulk

What are the intended semantic similarities?

- All are Male
  - but, so are 2 million other people
- All are Hollywood actors
  - but, so are 1.5 million other people
- All are Funny Actors
  - very few actors are funny

- SQuID rejects irrelevant semantic context

Not all semantic similarities are intended

- Implicit context
- Humans can tell easily!
- SQuID captures implicit semantic context

Example Tuples

Adam Sandler
Eddie Murphy
Jim Carrey

Filters encode semantic properties and constitute selection predicates

- Real-time performance through precomputing basic and derived semantic properties

Architecture

- Offline: Precomputes derived semantic properties and related statistics for real-time performance
- Online: Discovers semantic context, captures most likely query intent, and constructs SQL query

Example tuples

Iron Man
Spider Man
The Incredible Hulk

Example tuples

Adam Sandler
Eddie Murphy
Jim Carrey

Result tuples

Ant Man
The Avengers
Iron Man
The Incredible Hulk
Thor: The Dark World
Doctor Strange
Captain America
Spider Man
Guardians of the Galaxy

Example tuples

Iron Man
Spider Man
The Incredible Hulk

Example tuples

Adam Sandler
Eddie Murphy
Jim Carrey

Example tuples

Iron Man
Spider Man
The Incredible Hulk

Example tuples

Adam Sandler
Eddie Murphy
Jim Carrey