Facilitating Entity Navigation through Top-K Link Patterns

Liang Zheng, Yuzhong Qu, Jidong Jiang, Gong Cheng
Nanjing University, China
Entity Navigation

- **Entity navigation** over Linked Data often follows **semantic links** by using Linked Data browsers.
- **Semantic links** enable users to navigate between different entities.
Challenges

• Large numbers of linked entities and high diversity of links among entities.

About: **Steven Spielberg**

An Entity of Type: `yago:SpecialEffectsPeople`, within Data Space
Type: `yago:SpecialEffectsPeople`

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>is</code> producer of</td>
<td><code>The Hundred-Foot Journey</code></td>
</tr>
<tr>
<td></td>
<td><code>The Land Before Time</code></td>
</tr>
<tr>
<td></td>
<td><code>Men in Black</code></td>
</tr>
<tr>
<td></td>
<td><code>Poltergeist (Film)</code></td>
</tr>
<tr>
<td></td>
<td><code>Into the West (miniseries)</code></td>
</tr>
<tr>
<td></td>
<td><code>Amnesiac</code></td>
</tr>
<tr>
<td></td>
<td><code>Jessica Capshaw</code></td>
</tr>
<tr>
<td></td>
<td><code>Amy Irving</code></td>
</tr>
<tr>
<td></td>
<td><code>Your Studio and You</code></td>
</tr>
<tr>
<td><code>is</code> relatives of</td>
<td><code>Chambre 666</code></td>
</tr>
<tr>
<td></td>
<td><code>The Cutting Edge: The Magic of Movie Editing</code></td>
</tr>
<tr>
<td></td>
<td><code>Directed by John Ford</code></td>
</tr>
<tr>
<td></td>
<td><code>The Goonies</code></td>
</tr>
<tr>
<td></td>
<td><code>Ace Eli and Rodger of the Skies</code></td>
</tr>
<tr>
<td><code>is</code> spouse of</td>
<td><code>The Dig</code></td>
</tr>
<tr>
<td><code>is</code> starring of</td>
<td><code>Amblin’</code></td>
</tr>
<tr>
<td><code>is</code> story of</td>
<td><code>Medal of Honor: Allied Assault</code></td>
</tr>
<tr>
<td><code>is</code> writer of</td>
<td></td>
</tr>
</tbody>
</table>

51 semantic links

117 linked entities
Challenges

• Only depending on link traversal, users would have to browse a long list of semantic links, and search useful information by themselves. This procedure is often time-consuming.

How to improve the efficiency of entity navigation?
Sin

gle

‐

oriented

Entiti

‐

set

‐

oriented

• Potential relationships between the current entity (entities) and its (their) linked entities.
• Jurassic Park and A.I. were directed by Steven Spielberg.
• A.I. was directed and produced by Steven Spielberg.
• Dakota Fanning, Ben Kingsley and Morgan Freeman narrated at least 1 film in S.
• Tom Cruise and Dakota Fanning starred at least 2 films in S.
• Dakota Fanning narrated at least 1 film and also starred at least 2 films in S.
Link Pattern

• **Link pattern** represents the rich semantic relationships

  Definition 1 (Link Pattern with Minimum Number Restriction) Let $T$ be a link graph, $k$ be a positive integer, $l \in L$. A link pattern of $l$ with minimum $k$ restriction, denoted by $LP((\min k), l)$, is a function from $2^U$ to $2^U$ such that $LP((\min k), l)(S) = \{ v \in U | \{ u \in S | (u, l, v) \in T \} \geq k \}$ for $S \subseteq U$.

• **Jurassic Park** and A.I. were directed by **Steven Spielberg**.
  ✓ $(\min 1)\text{director}^{-1}(\text{Steven Spielberg}) = \{ \text{Jurassic Park, A.I.} \}$
  ✓ $\text{director}^{-1}(\text{Steven Spielberg}) = \{ \text{Jurassic Park, A.I.} \}$ //abbreviated

• **Dakota Fanning, Ben Kingsley** and **Morgan Freeman** narrated at least 1 film in $S$.
  ✓ $(\min 1)\text{narrator}(S) = \{ \text{Dakota Fanning, Ben Kingsley, Morgan Freeman} \}$
  ✓ $\text{narrator}(S) = \{ \text{Dakota Fanning, Ben Kingsley, Morgan Freeman} \}$ //abbreviated

• **Tom Cruise** and **Dakota Fanning** starred at least 2 films in $S$.
  ✓ $(\min 2)\text{starring}(S) = \{ \text{Tom Cruise, Dakota Fanning} \}$
Link Pattern

- **Link pattern** represents the rich semantic relationships

Definition 1 (Link Pattern with Minimum Number Restriction) Let $T$ be a link graph, $k$ be a positive integer, $l \in L$. A link pattern of $l$ with minimum $k$ restriction, denoted by $LP((\text{min } k), l)$, is a function from $2^U$ to $2^U$ such that

$$LP((\text{min } k), l)(S) = \{ v \in U \| \{ u \in S \| (u, l, v) \in T \} \geq k \} \text{ for } S \subseteq U.$$  

Definition 2 (Conjunctive Link Pattern) Given two link patterns $LP_1$ and $LP_2$, the conjunctive link pattern of $LP_1$ and $LP_2$, denoted by $LP_1 \land LP_2$, is a function from $2^U$ to $2^U$ such that

$$(LP_1 \land LP_2)(S) = LP_1(S) \cap LP_2(S) \text{ for } S \subseteq U.$$  

- **A.I.** was directed and produced by **Steven Spielberg**.

  ✓ $\text{director}^{-1} \land \text{producer}^{-1} (\text{Steven Spielberg}) = \{ \text{A.I.} \}$

- **Dakota Fanning** narrated at least 1 film and also starred at least 2 films in $S$.

  ✓ $(\text{narrator} \land ((\text{min } 2)\text{starring})) (S) = \{ \text{Dakota Fanning} \}$
Link Pattern

- **Link pattern** represents the rich semantic relationships

**Definition 1 (Link Pattern with Minimum Number Restriction)** Let $T$ be a link graph, $k$ be a positive integer, $l \in L$. A link pattern of $l$ with minimum $k$ restriction, denoted by $LP((\text{min } k), l)$, is a function from $2^U$ to $2^U$ such that

$$LP((\text{min } k), l)(S) = \{v \in U \mid \{u \in S \mid (u, l, v) \in T\} \geq k\} \text{ for } S \subseteq U.$$  

**Definition 2 (Conjunctive Link Pattern)** Given two link patterns $LP_1$ and $LP_2$, the conjunctive link pattern of $LP_1$ and $LP_2$, denoted by $LP_1 \land LP_2$, is a function from $2^U$ to $2^U$ such that

$$(LP_1 \land LP_2)(S) = LP_1(S) \cap LP_2(S) \text{ for } S \subseteq U.$$  

**Definition 3 (Sub-pattern Relationship)** Given two link patterns $LP_1$ and $LP_2$, $LP_1$ is called a sub-pattern of $LP_2$, denoted by $LP_1 \subseteq LP_2$, if $LP_1(S) \subseteq LP_2(S)$ holds for every subset $S$ of $U$.

- $\text{director}^{-1} \land \text{producer}^{-1} (\text{Steven Spielberg}) \subseteq \text{director}^{-1} (\text{Steven Spielberg})$
- $(\text{min } 2)\text{starring } (S) \subseteq \text{starring } (S)$
Link Pattern

- Link pattern represents the rich semantic relationships

Goal: To recommend $k$ link patterns for entity navigation
Step 1: Discover link patterns and sub-pattern relationships among them

A link pattern lattice is constructed based on Formal Concept Analysis

<table>
<thead>
<tr>
<th></th>
<th>starring (min2)starring narrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cruise</td>
<td>×</td>
</tr>
<tr>
<td>Fanning</td>
<td>× ×</td>
</tr>
<tr>
<td>Kingsley</td>
<td>× ×</td>
</tr>
<tr>
<td>Freeman</td>
<td>×</td>
</tr>
</tbody>
</table>

a: Constructing a formal context

b: Generating a link pattern lattice
Step 2: Select $k$ “good” link patterns

The “goodness” of link pattern contains:

- Informativeness
- Conciseness
- Specificness
Step 2: Select $k$ “good” link patterns

The “goodness” of link pattern contains:

- **Informativeness**

  ✓ A link pattern having fewer reachable linked entities is more informative.

  - $|\text{director}^{-1} \land \text{producer}^{-1} (\text{Steven Spielberg})| < |\text{director}^{-1} (\text{Steven Spielberg})|$

  - $\rightarrow \text{info}(\text{director}^{-1} \land \text{producer}^{-1} (\text{Steven Spielberg})) > \text{info}(\text{director}^{-1} (\text{Steven Spielberg}))$
Step 2: Select $k$ “good” link patterns

The “goodness” of link pattern contains:

- Conciseness

✓ A concise link pattern having a shorter label is more understandable and preferable

☐ $\text{label (narrator} \land ((\text{min 2)starring})) > \text{label ((min 2)starring})$

☐ $\rightarrow \text{conc(narrator} \land ((\text{min 2)starring})) < \text{conc((min 2)starring})$
Step 2: Select $k$ “good” link patterns

The “goodness” of link pattern contains:

- **Specificness**

  ✓ The larger the depth of link pattern is, the more specific the link pattern is.

  - $\text{depth } ((\text{min } 2)\text{starring}) > \text{depth } (\text{starring})$
  - $\rightarrow \text{spec } ((\text{min } 2)\text{starring}) < \text{spec } (\text{starring})$
Step 2: Select $k$ “good” link patterns

- For diversity and coverage considerations, selecting $k$ link patterns that
  - are as informative, concise and specific as possible
  - while being able to retrieve as many linked entities as possible.

- Formalized based on **Budgeted Maximum Coverage** problem.

- Solved by an approximation algorithm
Overview of Prototype

http://ws.nju.edu.cn/sview/
User Study

• 3 participant systems
  – SView
  – OpenLink Faceted Search & Find Service
  – Rhizomer

• 24 subjects

• 80 navigation tasks over DBpedia

<table>
<thead>
<tr>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G₁</strong></td>
</tr>
<tr>
<td><em>E₁</em> Explore the information related to Steven Spielberg.</td>
</tr>
<tr>
<td><em>F₁</em> Find the films directed and also produced by Steven Spielberg.</td>
</tr>
<tr>
<td><strong>G₂</strong></td>
</tr>
<tr>
<td><em>E₂</em> Explore the information related to the films directed by Steven Spielberg.</td>
</tr>
<tr>
<td><em>F₂</em> Find the actors starred in at least 2 films directed by Steven Spielberg.</td>
</tr>
</tbody>
</table>
# Navigation Questionnaire Results

## Questions

- **Q1:** The number of navigation options (links) was overwhelming.
- **Q2:** The navigation options (links) were well organized.
- **Q3:** The navigation option (link) titles were understood well.
- **Q4:** The navigation options (links) were pleasantly surprising.
- **Q5:** It was easy to reorient myself in the navigation.

## Results

<table>
<thead>
<tr>
<th>Questions</th>
<th>Response: Mean (SD)</th>
<th>F(2, 69)</th>
<th>LSD post-hoc (p &lt; 0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OpenLink</td>
<td>Rhizomer</td>
<td>SView</td>
</tr>
<tr>
<td>Q1:</td>
<td>3.919</td>
<td>3.75</td>
<td>2.667</td>
</tr>
<tr>
<td></td>
<td>(0.717)</td>
<td>(0.854)</td>
<td>(1.095)</td>
</tr>
<tr>
<td>Q2:</td>
<td>3.026</td>
<td>3.24</td>
<td>4.11</td>
</tr>
<tr>
<td></td>
<td>(1.052)</td>
<td>(1.014)</td>
<td>(0.887)</td>
</tr>
<tr>
<td>Q3:</td>
<td>3.833</td>
<td>4.00</td>
<td>2.583</td>
</tr>
<tr>
<td></td>
<td>(0.717)</td>
<td>(0.582)</td>
<td>(0.62)</td>
</tr>
<tr>
<td>Q4:</td>
<td>2.58</td>
<td>3.25</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>(0.793)</td>
<td>(1.055)</td>
<td>(0.778)</td>
</tr>
<tr>
<td>Q5:</td>
<td>3.917</td>
<td>3.667</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>(0.514)</td>
<td>(0.778)</td>
<td>(0.937)</td>
</tr>
</tbody>
</table>
Performance Evaluation

$m$: the number of current entities
Performance Evaluation

$m$: the number of current entities
Conclusion and Future Work

- Proposed link pattern
- Constructed link pattern lattice
- Recommended Top-k link patterns

- Future work
  - Entity Type
  - Organize semantic links and linked entity types simultaneously.
THANK YOU!
QUESTIONS?