Identifying Potentially Important Concepts and Relations in an Ontology

Gang Wu, Juanzi Li, Ling Feng, and Kehong Wang
wug@keg.cs.tsinghua.edu.cn, wu@seu.edu.cn

The 7th International Semantic Web Conference
October 28, 2008
1 Introduction

2 CARRank
   - CARRank Model
   - CARRank Algorithm

3 Experiments
   - Experimental Settings
   - Ranking Qualities
   - Comparison of Semantic Abilities
   - Efficiency

4 Conclusion & Discussion
Motivation

- Amount, Scale, and Complexity of ontologies
- Efforts of understanding an ontology
Motivation

- Amount, Scale, and Complexity of ontologies
- Efforts of understanding an ontology
- Helps from computer intelligence
  - Information visualization: IsaViz, Ontoviz, and Jambalaya
Motivation

- Amount, Scale, and Complexity of ontologies
- Efforts of understanding an ontology
- Helps from computer intelligence
  - Information visualization: IsaViz, Ontoviz, and Jambalaya
Motivation

- Amount, Scale, and Complexity of ontologies
- Efforts of understanding an ontology
- Helps from computer intelligence
  - Information visualization: IsaViz, Ontoviz, and Jambalaya ...

- Drawing users’ attention to important concepts
  - Browsing activities: DIaMOND project [dS06]
  - Concept hierarchy [TXZ+05]
Motivation

- **Amount, Scale, and Complexity of ontologies**
- **Efforts of understanding an ontology**
- **Helps from computer intelligence**
  - Information visualization: IsaViz, Ontoviz, and Jambalaya ...

- Drawing users’ attention to important concepts
  - Browsing activities: DIaMOND project [dS06]
  - Concept hierarchy [TXZ⁺05]
Our Approach

- Help understanding an ontology
  - With more ontology information (structure), and easier to use
Our Approach

- Help understanding an ontology
  - With more ontology information (structure), and easier to use

- Concept And Relation Ranking (CARRank)
  - Identifying potentially important concepts and relations in an ontology simultaneously
  - Importance: the degree the creator would like to suggest
  - Conducting complex networks analysis on ontology graph
Cognitive Support for Ontology Understanding

- **DIaMOND project**: a plug-in for Protégé [dS06]
- In [TXZ+05]: an Eclipse application
Related Work

- Cognitive Support for Ontology Understanding
  - DIAmond project: a plug-in for Protégé [dS06]
  - In [TXZ+05]: an Eclipse application

- Link Analysis Ranking algorithms
  - PageRank [BP98], HITS [Kle99], and Reverse PageRank [Fog03]
  - ObjectRank [BHP04], and PopRank [NZWM05]
Related Work

- Cognitive Support for Ontology Understanding
  - DIAmond project: a plug-in for Protégé [DS06]
  - In [TXZ+05]: an Eclipse application

- Link Analysis Ranking algorithms
  - PageRank [BP98], HITS [Kle99], and Reverse PageRank [Fog03]
  - ObjectRank [BHP04], and PopRank [NZWM05]

- Ontology Ranking in the Semantic Web
  - OntoSelect [BED04], and OntoKhoj [PSLP03]
  - AKTiveRank [ABS06]
Outline

1. Introduction

2. CARRank
   - CARRank Model
   - CARRank Algorithm

3. Experiments
   - Experimental Settings
   - Ranking Qualities
   - Comparison of Semantic Abilities
   - Efficiency

4. Conclusion & Discussion
Ontology Graph

Definition

Given an ontology \( \mathcal{O} \), the **ontography graph** \( G = (\mathcal{V}, \mathcal{E}, l_\mathcal{V}, l_\mathcal{E}) \) of \( \mathcal{O} \) is a directed labeled graph. \( \mathcal{V} \) is a set of nodes representing all concepts in \( \mathcal{O} \). \( \mathcal{E} \) is a set of directed edges representing all relations in \( \mathcal{O} \). \( l_\mathcal{V} \) and \( l_\mathcal{E} \) are labeling functions on \( \mathcal{V} \) and \( \mathcal{E} \) respectively.
Ontology Graph Example
From RDF Graph to Ontology Graph

RDF Graph $\neq$ Ontology Graph

- Triples
  
  $(\text{manage, rdfs:domain, Project_Admin})$
  
  and $(\text{manage, rdfs:range, Project})$

- RDF Graph
  
  [Diagram showing RDF Graph with nodes labeled Project_Admin and Project, and a manage edge with rdfs:domain and rdfs:range properties]

- Ontology Graph
  
  [Diagram showing Ontology Graph with a manage edge between Project_Admin and Project]
Mapping

Definition

Let $G = (V, E, l_V, l_E)$ be the RDF graph of an ontology $\mathcal{O}$. We define a map $\omega : G \rightarrow G$ as follows: $\omega(G) = (\mathcal{V}, \mathcal{E}, l_\mathcal{V}, l_\mathcal{E})$ where,

- $\mathcal{V} = V$, $l_\mathcal{V} = l_V$,
- $\mathcal{E} = \{ e_{s,p,o} | e_{s,p,o} \in E \land l_E(e_{s,p,o}) \neq \text{rdfs:domain} \land l_E(e_{s,p,o}) \neq \text{rdfs:range} \}$

  $\cup E_{DR} \cup E_D \cup E_R$,

- $E_{DR} = \{ e_{s,p,o} | \exists e_p, \text{rdfs:domain}, s \in E \land \exists e_p, \text{rdfs:range}, o \in E \}$,

- $E_D = \{ e_{s,p,\text{keg:Sink}} | \exists e_p, \text{rdfs:domain}, s \in E \land \forall e_p, \text{rdfs:range}, o \in E \}$,

- $E_R = \{ e_{\text{keg:Source},p,o} | \exists e_p, \text{rdfs:range}, o \in E \land \forall e_p, \text{rdfs:domain}, s \in E \}$,

- $\forall e_{s,p,o} \in \mathcal{E}$, from$(e_{s,p,o}) = v_s$, to$(e_{s,p,o}) = v_o$, and $l_\mathcal{E}(e_{s,p,o}) = p$

Here, \text{keg:Source} and \text{keg:Sink} are defined to be the virtual domain and range of those relations having no domain or range defined explicitly.
Model Description

- Four features for potentially important concepts and relations
  - A concept is more important if there are more relations starting from the concept
  - A concept is more important if there is a relation starting from the concept to a more important concept
  - A concept is more important if it has a higher relation weight to any other concept
  - A relation weight is higher if it starts from a more important concept

- Note
  - A concept is regarded as a hub
  - Concepts and relations mutually reinforce
Semantic Abilities

- For any RDF-based ontology: RDF Schema, and OWL
- To support axioms expressed as rules: e.g. SWRL
- Only considering the local importance for relations
- Ontology understanding is affected by many factors
Semantic Abilities

- For any RDF-based ontology: RDF Schema, and OWL
- To support axioms expressed as rules: e.g. SWRL
- Only considering the local importance for relations
- Ontology understanding is affected by many factors
Semantic Abilities

- For any RDF-based ontology: RDF Schema, and OWL
- To support axioms expressed as rules: e.g. SWRL
- Only considering the local importance for relations
- Ontology understanding is affected by many factors
Semantic Abilities

- For any RDF-based ontology: RDF Schema, and OWL
- To support axioms expressed as rules: e.g. SWRL
- Only considering the local importance for relations
- Ontology understanding is affected by many factors
Outline

1. Introduction
2. CARRank
   - CARRank Model
   - CARRank Algorithm
3. Experiments
   - Experimental Settings
   - Ranking Qualities
   - Comparison of Semantic Abilities
   - Efficiency
4. Conclusion & Discussion
## CARRank Algorithm

### Equation

1. **Equation 1:**
   
   \[ w_{k+1}(s, t) = \frac{r_k(s)}{\sum_{t_i \in B_t} r_k(t_i)} \]

2. **Equation 2:**
   
   \[ r_{k+1}(s) = \frac{1 - \alpha}{|\mathcal{V}|} + \alpha \sum_{t_i \in F_s} r_k(t_i) w_{k+1}(s, t_i) \]

- \( r_{k+1}(s) \) and \( w_{k+1}(s, t) \): in the \( k + 1 \) step, the importance of a concept \( s \in \mathcal{V} \) and the weight of relation from \( s \) to another concept \( t \in \mathcal{V} \)

- **Concept importance vector:** \( \mathbf{R} = (r_1, ..., r_n) \)

- Concept \( v_i \)'s relation importance vector:
  
  \( \mathbf{L}_i = (r_1 w_{i,1}, \ldots, r_n w_{i,n}) \)

- The iterative process stop at \( \| \mathbf{R}_{k+1} - \mathbf{R}_k \| < \varepsilon \)
1 Introduction

2 CARRank
   - CARRank Model
   - CARRank Algorithm

3 Experiments
   - Experimental Settings
   - Ranking Qualities
   - Comparison of Semantic Abilities
   - Efficiency

4 Conclusion & Discussion
Experimental Settings

- **Evaluation Metrics**
  - Convergence efficiency: the number of iterations $k$ minimizing $\| R_{k+1} - R_k \|$ to a given threshold $\varepsilon$
  - Ranking quality
    - Concept: $\widehat{P@20} = \frac{n_1 \sim 3 \times 20 + n_4 \sim 10 \times 17 + n_{11} \sim 20 \times 10}{279}$, [LS99]
    - Relation: $PR = \frac{\sum_{c \in C_{1 \sim 20}} m_c}{|C_{1 \sim 20}|}$

- **Ranking Methods**
  - Standard PageRank (PR) [BP98]
  - Concept hierarchy (CH) [TXZ+05]
  - AKTiveRank’s density and betweenness measures (DEM+BEM) [ABS06]

- **Experimental Environments:** $\alpha = 0.85, \varepsilon = 1 \times 10^{-6}$
Outline

1. Introduction
2. CARRank
   - CARRank Model
   - CARRank Algorithm
3. Experiments
   - Experimental Settings
   - Ranking Qualities
   - Comparison of Semantic Abilities
   - Efficiency
4. Conclusion & Discussion
**Table:** Four ontologies

<table>
<thead>
<tr>
<th>Concept#</th>
<th>Property#</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWL</td>
<td>17</td>
<td><a href="http://www.w3.org/2002/07/owl.rdf">http://www.w3.org/2002/07/owl.rdf</a></td>
</tr>
<tr>
<td>Software Project</td>
<td>14</td>
<td><a href="http://keg.cs.tsinghua.edu.cn/persons/tj/ontology/software.owl">http://keg.cs.tsinghua.edu.cn/persons/tj/ontology/software.owl</a></td>
</tr>
<tr>
<td>Copyright Ontology</td>
<td>98</td>
<td><a href="http://rhizomik.net/ontologies/2006/01/copyrightonto.owl">http://rhizomik.net/ontologies/2006/01/copyrightonto.owl</a></td>
</tr>
<tr>
<td>Travel Ontology</td>
<td>84</td>
<td><a href="http://learn.tsinghua.edu.cn:8080/2003214945/travelontology.owl">http://learn.tsinghua.edu.cn:8080/2003214945/travelontology.owl</a></td>
</tr>
</tbody>
</table>

- **Ranking task:**
  
  *For each ontology file, list top 20 (or as many as you like) important concepts (with URI) of your ontology in your mind. And for each top concept, please give top 5 (or as many as you like) important relations (with URI) for that concept.*

- **Reference answers:** ontology creators’ feedback to the task
- **Inquiry method:** email to the four contact creators
User study:
- Five volunteers who interest in the Semantic Web
- Unfamiliar with the ontologies before
- Everyone independently gave the top 20 important concepts and the top 5 important relations for each top concept
- Computed a $\hat{P}@20$ value and a $PR$ value
- Arithmetic means on five $\hat{P}@20$ values and five $PR$ values
The importance of concepts – Software ontology

Table: The importance of concepts – Software ontology

<table>
<thead>
<tr>
<th>Rank</th>
<th>Reference Answer</th>
<th>PageRank</th>
<th>DEM+BEM</th>
<th>CARRank</th>
<th>User Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Project</td>
<td>Message</td>
<td>Project</td>
<td>Project</td>
<td>Project</td>
</tr>
<tr>
<td>2</td>
<td>Member</td>
<td>has_usage_statistics</td>
<td>Usage_statistics</td>
<td>Usage_statistics</td>
<td>Category</td>
</tr>
<tr>
<td>3</td>
<td>Developer</td>
<td>statistics_bugs</td>
<td>Developer</td>
<td>Statistic_record</td>
<td>Message</td>
</tr>
<tr>
<td>4</td>
<td>Category</td>
<td>statistic_record_support</td>
<td>Statistic_record</td>
<td>Developer</td>
<td>Discussion</td>
</tr>
<tr>
<td>5</td>
<td>Public_forum</td>
<td>Member</td>
<td>Member</td>
<td>Category</td>
<td>Help</td>
</tr>
<tr>
<td>6</td>
<td>LatestNew</td>
<td>Project</td>
<td>Message</td>
<td>Release_package</td>
<td>Person</td>
</tr>
<tr>
<td>7</td>
<td>Message</td>
<td>Developer</td>
<td>Public_forums</td>
<td>Member</td>
<td>Member</td>
</tr>
<tr>
<td>8</td>
<td>Version</td>
<td>Category</td>
<td>Person</td>
<td>Message</td>
<td>Developer</td>
</tr>
<tr>
<td>9</td>
<td>homepage</td>
<td>super_category</td>
<td>Category</td>
<td>Help</td>
<td>Project_admin</td>
</tr>
<tr>
<td>10</td>
<td>Usage_statistics</td>
<td>page_views</td>
<td>Project_admin</td>
<td>Public_forums</td>
<td>Public_forums</td>
</tr>
</tbody>
</table>
## The importance of relations – Software ontology

**Table:** The importance of relations – Software ontology

<table>
<thead>
<tr>
<th>Top 5 Concepts</th>
<th>Reference Answer</th>
<th>Ranking results</th>
<th>User Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>title</td>
<td>has_usage_statistics</td>
<td>project_homepage</td>
</tr>
<tr>
<td>2</td>
<td>summary</td>
<td>developed_by</td>
<td>title</td>
</tr>
<tr>
<td>3</td>
<td>activity_ranking</td>
<td>belong_to_category</td>
<td>activity_ranking</td>
</tr>
<tr>
<td>4</td>
<td>project_homepage</td>
<td>translations</td>
<td>has_public_forum</td>
</tr>
<tr>
<td>5</td>
<td>project_of_statistic</td>
<td>intended_audience</td>
<td>has_usage_statistics</td>
</tr>
<tr>
<td><strong>Member</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>login_name</td>
<td>post_message</td>
<td>person_name</td>
</tr>
<tr>
<td>2</td>
<td>publicly_displayed_name</td>
<td>site_member_since</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>email_address</td>
<td>login_name</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>user_id</td>
<td>email_address</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>site_member_since</td>
<td>publicly_displayed_name</td>
<td></td>
</tr>
<tr>
<td><strong>Developer</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>skills</td>
<td>member_of_project</td>
<td>person_name</td>
</tr>
<tr>
<td>2</td>
<td>project_role</td>
<td>project_role</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>skills</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>user_id</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Category</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>hasProject</td>
<td>hasProject</td>
<td>super_category</td>
</tr>
<tr>
<td>2</td>
<td>category_name</td>
<td>sub_category</td>
<td>sub_category</td>
</tr>
<tr>
<td>3</td>
<td>super_category</td>
<td>super_category</td>
<td>category_name</td>
</tr>
<tr>
<td>4</td>
<td>sub_category</td>
<td>category_name</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Public_Forum</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>hasMessage</td>
<td>hasMessage</td>
<td>hasMessage</td>
</tr>
<tr>
<td>2</td>
<td>belong_to_project</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>project_of_forum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Comparison of Ranking Concepts & Relations

Figure: The Comparison of Ranking Concepts

Table: The Comparison of Ranking Relations

<table>
<thead>
<tr>
<th>Ontology</th>
<th>CARRank</th>
<th>User Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software</td>
<td>0.586</td>
<td>0.562</td>
</tr>
<tr>
<td>Copyright</td>
<td>0.06</td>
<td>0</td>
</tr>
</tbody>
</table>
Outline

1. Introduction

2. CARRank
   - CARRank Model
   - CARRank Algorithm

3. Experiments
   - Experimental Settings
   - Ranking Qualities
   - Comparison of Semantic Abilities
   - Efficiency

4. Conclusion & Discussion
For Languages with Different Expressive Abilities

We generated 3 variations of FOAF ontology, i.e. OWL-Full, OWL-DL, and OWL-Lite, with a tool named foaf_cleaner [Alf05]

<table>
<thead>
<tr>
<th>Table: Top Concepts for FOAF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Person</td>
</tr>
<tr>
<td>Document</td>
</tr>
<tr>
<td>Organization</td>
</tr>
<tr>
<td>Project</td>
</tr>
<tr>
<td>Agent</td>
</tr>
<tr>
<td>OnlineEcommerceAccount</td>
</tr>
<tr>
<td>OnlineChatAccount</td>
</tr>
<tr>
<td>OnlineGamingAccount</td>
</tr>
<tr>
<td>OnlineAccount</td>
</tr>
<tr>
<td>PersonalProfileDocument</td>
</tr>
<tr>
<td>Image</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Pearson Correlation Coefficients</td>
</tr>
</tbody>
</table>
1 Introduction

2 CARRank
   - CARRank Model
   - CARRank Algorithm

3 Experiments
   - Experimental Settings
   - Ranking Qualities
   - Comparison of Semantic Abilities
   - Efficiency

4 Conclusion & Discussion
Convergence Comparison

- **Methods**: PageRank, Reverse PageRank, and CARRank
- **Two ontologies**
  - “Relationship”: 169 vertices and 252 directed labeled edges
  - “UNSPSC”: 19600 vertices and 29386 directed labeled edges

![Graphs showing convergence comparison for different methods.](image-url)
Conclusion

- CARRank is a simple yet effective algorithm
- Identifying potentially important concepts and relations in an ontology
- The experimental results show the feasibility

Discussion

- Ontology understanding means much more
- User-independent vs. Based on users’ tasks & needs
- CARRank is a preliminary step, may be not the best
Thank You

Q & A
Harith Alani, Christopher Brewster, and Nigel Shadbolt.
Ranking ontologies with aktiverank.
In ISWC, Athens, GA, USA, November 2006.

Ron Alford.
Using FOAF and OWL, July 2005.

Paul Buitelaar, Thomas Eigner, and Thierry Declerck.
Ontoselect: A dynamic ontology library with support for ontology selection.
In The Demo Session at the ISWC, 2004.

Andrey Balmin, Vagelis Hristidis, and Yannis Papakonstantinou.
Objectrank: Authority-based keyword search in databases.

Sergey Brin and Lawrence Page.
The anatomy of a large-scale hypertextual web search engine.

Tricia d’Entremont and Margaret-Anne Storey.
Using a degree-of-interest model for adaptive visualizations in protégé.
In 9th International Protégé Conference, 2006.

Dániel Fogaras.
Where to start browsing the web?
Jon M. Kleinberg.
Authoritative sources in a hyperlinked environment.

H. Vernon Leighton and Jaideep Srivastava.
First 20 precision among world wide web search services (search engines).

Zaiqing Nie, Yuanzhi Zhang, Ji-Rong Wen, and Wei-Ying Ma.
Object-level ranking: bringing order to web objects.

Ontokhoj: a semantic web portal for ontology searching, ranking and classification.

Kewei Tu, Miao Xiong, Lei Zhang, Haiping Zhu, Jie Zhang, and Yong Yu.
Towards imaging large-scale ontologies for quick understanding and analysis.
In *ISWC*, 2005.