

Defining Semantic Relationships to Capitalize Content of Multimedia Resources

Mohamed Kharrat^(✉), Anis Jedidi, and Faiez Gargouri

MIRACL - Multimedia, InfoRmation systems and Advanced Computing Laboratory,
University of Sfax, Sfax, Tunisia
med_khr@yahoo.fr, anis.jedidi@isimsf.rnu.tn,
faiez.gargouri@isimsf.rnu.tn

Abstract. Existing systems or architectures hardly provide any way to localize sub-parts of multimedia objects (e.g. sub regions of images, persons, events...), which represents hidden semantics of resources. To simplify and automate discovering hidden connections between such resources, we describe and evaluate in this paper, an algorithm for creating semantic relationships between multimedia news resources, giving a contextual schema (represented in RDF) as a result. This latter, which could eventually be used under any retrieval system, is integrated in our main multimodal retrieval system.

We have also proposed and introduced a special measure of accuracy since evaluation relies on users' intentions. An experimental evaluation of our algorithm is presented, showing encouraging results.

Keywords: Semantic · Multimedia · Relationships

1 Introduction

Semantic web is one of the most important challenges in web realm which has been subject of many researches in recent years.

One of the main elements for processing semantic web is to go further knowledge and annotation. In fact, developing semantic retrieval systems, needs information extraction, harvesting knowledge and various methods of data.

Advances in multimedia technologies have made possible the storage of huge multimedia documents collections on computer systems. But the lack of efficiency is perceived as the main obstacle for a large-scale deployment of semantic technologies. In order to allow an efficient exploitation of these collections, designing tools for accessing data resources is required.

One of the biggest challenges is the exploitation of these collections, particularly hidden or non-exploitable relations as well as search and querying. To address this problem, we propose a mechanism to generate a new defined set of hidden semantic relationships between multimedia documents.

Within the same framework, our main system [5] proposes to retrieve multimedia documents using a multimodal approach. The main characteristic of our system is the

use of two languages, XQuery and SPARQL to query the description of multimedia resources. Performance of our system can be significantly increased by using a semantic relationship contextual schema (see Figure1) for semantic relationships, by applying rules through an algorithm described in this paper.

The importance of discovering such links is essentially for retrieving relevant hidden resources in results.

The algorithm which is exposed here, allows the generation and publication of linked data from metadata. Any resource which is composed of many parts could or could not have many relationships with other resources.

In addition, we introduce a special measure that allows a user to rate the correctness of each relationship and penalize irrelevant ones based on its own perception.

The aim of this paper is to present how to implement semantic relationships between data along with multimedia news resources to enhance our ability to "understand" those latter ones. In fact, news descriptive meta-data available for users, are difficult to learn about their content and capabilities, this is why we are seeking for strengthening by establishing this feature to our main system.

The next step will be the integration of this mechanism over XQuery language, which gives the possibility to add new relationships through queries. This means, we will create new relationships based on the resources which are the results of queries. In fact, XQuery will be first used to build semantic relationships over queries based on functions, and secondly, it will be used to harness them .

The last part of the paper is structured as follows. Section 2 provides an overview of closely related work. We present proposed semantic relationships and a complementary inference reasoning to build these relationships in section 3. We evaluate our approach in section 4 and finally we conclude in section5.

2 Related Work

As far as we know, there are no other works reported addressing the task of creating semantic relationships from XML content of multimedia resources. Then, we are going to briefly present here some previous studies which are quite close to our work, consisting mainly on automatic identification of relationships from unstructured documents or the use of lexical patterns for relations discovery between concepts. These have the advantage of the simplicity of collecting training corpora automatically.

While authors use a graph in [6] to model relationships between phrases inside semantic corpus Wordnet using numbers, our method does not use a graph.

Authors in [7] Establish missing semantic relations between Wikipedia entities by discovering automatically the missing entity links in Wikipedia infoboxes, which are important for creating RDF links between DBpedia instances.

Several other approaches have been proposed having various methods and techniques. Solution proposed in [3], identify semantic links between persons, products, events and other entities from Twitter based on entities topics and their types according to time axis.

In [2], authors compute concept-concept relatedness and concept-category relatedness based on heuristics by Category links and related links in Wikipedia.

In [4], author proposes RelFinder: an interactive discovery of relationships between DBpedia objects which is controlled by users, combined with the automatic mechanisms according to topological and semantic dimensions.

Our problem shares some resemblance with works in [1], where author creates new links in precise region on images. This region represents the most relevant part in XML document of each image using hierarchical structure and adds weights for every link. The goal is to ameliorate the image retrieval in the semi-structured documents.

In contrast to our approach, all these works treat mono-media documents. They propose descriptions which allow establishing relations between annotated concepts, resources and parts of resources. These works do not take into account multimedia resources and the set of sturdy relations which are between resources or between parts of a same resource. In addition, most of them, do not consider the semantic side.

In the following section, we introduce relations and rules which allow us to extract semantics from multimedia resources.

3 Semantic Relationships

We introduce here, a contextual schema which constitutes formalism for semantic relationships representation. It expresses meaning in a form that is both logically precise and humanly readable. This schema is implemented to be used in our multimodal system and represented using RDF.

The basic assumption underlying our approach is textual descriptions of resources always hide semantics that cannot always be discovered notably between concepts. Besides, meaning of some data is sometimes either unknown, ambiguous or implicit.

However, not all semantically related concepts are interesting for end users. In this paper, we have identified a number of semantic relations.

Media fragments are really parts of a parent resource. The use of identifiers seems therefore appropriate to specify these media fragments. As for any identifier, access to the parent identifier shall be possible in order to inspect its context. Providing a way being used as agreed to localize sub-parts of multimedia objects (e.g. sub regions of images, temporal sequences of videos etc.) is fundamental.

All resources in our collection are described with NewsML annotation standard for news documents. In this standard, metadata itself comes in bewildering variety. There are specific terms to describe every type of media. We harness them to extract contextual relations to be used in semantic and contextual recognition. Most visual and audio features (motion, speech, text) will be used to describe each part. For example, in order to describe the content of video news, we apply concepts to describe scenes like meeting, speech, interview, live reporting or events/topics like sports, politics and commercials. Notably, we also apply the identities of persons that can be recovered from the visual flow (person who appears on the screen), from audio or from textual information.

Our goal is to make a semantic search based on both content and structure at the same time. We do not propose to use existing links between resources, but we create our own links.

Our algorithm takes as input a resource and generates a new relationship if links exit with some other resources.

```

<?xml version="1.0"?>
<resources><resource id="IMAGE01" type="image">
<link name= "AI">
<resource id="IMAGE02" type="image"></resource>
<resource id="VIDEO01" type="video"></resource>
</link>
</resources>
<resource id="VIDEO07" type="video"> <link name="SH"><resource
id="IMAGE03" type="image"></resource>
...
</link>

```

Fig. 1. Sample of Contextual Schema

3.1 Relationships Mechanism

The task of relationships building is a crossing problem between textual relations and semantic relations.

For instance, the textual expression “P talks in R” indicates a semantic relation Talk between entity “P” which is represented by a resource and another resource “R”.

This semantic relationship can be expressed textually in several ways: for example “P, said something about X” or “a quotation of P in R”.

There are several components to make a coherent relationship, including specific textual expressions as well as constraints on the entities involved in the relation. For instance, in the Talk relation, “P” must be a Person and “R” a Resource. The details of every relationship are given below.

T: Talk

This type of relationships describes links between resource R which contains {person, organization, team...} talking. This relation must be between an image and another type of document.

TA: Talk About

This type of relationships describes links between resource R which represents {document, report, documentary...} and another resource R'.

S: Speak

This type of relationships describes links between resource R which contains only a person and another resource R'.

SA: Speak About

This type of relationships describes links between resource R which contains only a person speaking, and another resource R'.

SH: Show

This type of relationships describes links between a resource R {documentary, event, interview...} which shows {person, organization, team, place...} and another resource R'.

AI: Appear In

This type of relationships describes links between a resource R which represents {person, organization, team, place...} and appears in another resource R' which represents {event, scene, sequence...}.

In the following, we briefly explain the mechanism via rules that must be used to create these relationships.

Algorithm

```

Input: Xml resource r
Output: relation between two or more resources in contextual schema CS
For all r ∈ {R} do
{Extract metadata from r and r'}
If any verified module
{ If in CS
Add new relation to CS
End If}
Else
Execute module inference
End If
End For
Return r ↔ r'

```

Rule 1:

$$\text{Talk}(\exists R \supset \{\text{image}\} \wedge R' \supset \{\text{video, audio, text}\} \wedge \exists \langle \text{object}, \langle \text{person}, \text{or} \langle \text{ganization} \rangle \rangle \supset R \supset R' \wedge \exists \{\langle \text{interview}, \langle \text{report} \rangle\} \supset R \quad (1)$$

To add the new relationship *Talk*, the resource origin R must be an image and the destination R' could be any type (video, audio, text). Secondly metadata like $\langle \text{object}, \langle \text{person} \rangle$ or $\langle \text{organization} \rangle$ must exist in the two resources. In addition, $\langle \text{interview} \rangle$ or $\langle \text{report} \rangle$ must be present in the destination resource.

Rule 2:

$$\text{Speak}(\exists R \supset \{\text{image}\} \wedge R' \supset \{\text{video, audio}\} \wedge \exists \{\langle \text{person} \rangle\} \supset R \supset R' \wedge \exists \{\langle \text{interview} \rangle \vee \langle \text{speech} \rangle\} \supset R' \quad (2)$$

To add the new relationship *Speak*, the resource origin R must be an image and the destination R' could be (video or audio). Secondly, the metadata only $\langle \text{person} \rangle$ must exist in the two resources. In addition, $\langle \text{interview} \rangle$ or $\langle \text{speech} \rangle$ must be present in the destination resource.

Rule 3:

$$\text{TalkAbout}(\exists R, R' \supset \{\text{image, video, audio, text}\} \wedge R \equiv R' \wedge \text{type}(R) \neq \text{type}(R') \quad (3)$$

To add the new relationship TalkAbout, the resource origin R and the destination R' could be of any type of media (image, video, audio, text). Secondly, we fix a similarity threshold between metadata of both resources using TFIDF measure between XML's tags of these resources. The type of related resources must be different, (e.g. we could not relate two images or two videos).

Rule 4:

$$\text{SpeakAbout}(\exists R \supset \{\text{video}, \text{audio}\} \wedge R' \supset \{\text{image}, \text{video}, \text{audio}, \text{text}\} \wedge \exists \langle \text{person} \rangle \supset R \wedge R \equiv R') \quad (4)$$

To add the new relationship SpeakAbout, the resource origin R must be a video or an audio resource and the destination R' could be of any type (image, video, audio, text). Secondly, metadata $\langle \text{person} \rangle$ must exist in the original resource. Finally, we fix a similarity threshold between metadata of both resources using TFIDF measure between XML tags of resources.

Rule 5:

$$\text{Show}(\exists R \supset \{\text{video}, \text{image}\} \wedge R' \supset \{\text{image}, \text{video}, \text{audio}, \text{text}\} \wedge \exists \{\langle \text{documentary} \rangle, \langle \text{event} \rangle, \langle \text{interview} \rangle\} \supset R \wedge R \equiv R') \quad (5)$$

To add the new relationship Show, the resource origin R could be only a video or an image and the destination R' could be any type (image, video, audio, and text). Secondly metadata {documentary, event, interview...} must exist in the original resource. Finally, we fix a similarity threshold between metadata of both resources using TFIDF measure between XML tags of resources.

Rule 6:

$$\text{AppearIn}(\exists R \supset \{\text{image}\} \wedge R' \supset \{\text{video}, \text{image}\} \wedge \exists \{\langle \text{object} \rangle, \langle \text{person} \rangle, \langle \text{organization} \rangle\} \supset R \supset R') \quad (6)$$

To add the new relationship AppearIn, the resource origin R must be only an image and the destination R' could be (image or video). Secondly metadata like $\langle \text{object} \rangle$, $\langle \text{person} \rangle$ or $\langle \text{organization} \rangle$ must exist in the two resources.

3.2 Inference reasoning

Since XML does not support or suggest reasoning mechanisms, we have to rely on an underlying logical formalism.

We define here some inductive rules to deduce new relationships from existing relationships.

$$\text{Case 1:} \quad \exists \text{link}(R1, R2) \wedge \text{proximity}(R3, R2) \Rightarrow \text{link}(R1, R3) \quad (7)$$

R1 \rightarrow R2 are two related resources, so if the new resource R3 has proximity with R2 then R1 \rightarrow R3

$$\text{Case 2:} \quad \exists (\text{link}(R1, \{R\}) \equiv \text{link}(R2, \{R\})) \Rightarrow R1 = R2 \quad (8)$$

$$\text{Case 3:} \quad \exists (\text{link}(R1, R2) \wedge \text{link}(R1, R2) \wedge \text{link}(R1, R3)) \Rightarrow R2 \equiv R3 \equiv R4 \quad (9)$$

If there is semantic relationship between resource R1 and other resources as follow:

$R1 \rightarrow R2 ; R1 \rightarrow R3 ; R1 \rightarrow R4$

Then, there will be similarities between R2, R3 and R4.

We define link (R,R') as an existing semantic relation between R and R' and Proximity (R,R') as the similarity between R and R' calculated by the measure below.

3.3 Similarity Measure

We use this computation whenever a similarity measure is needed. It is composed of three steps.

First step:

Pre-processing: this module is concerned with pre-processing operations preparing the input resource to be linked. It checks if a resource R is typed.

Second step:

- Comparing <keyword> of R and R' for equality or similarity
- Comparing <title> of R and R' for equality or similarity

Third step:

Similarity is defined by some functions:

The Jaccard coefficient measures similarity between sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (10)$$

In addition, we use term frequency. This count is usually normalized to prevent a bias towards longer documents (which may have a higher term count regardless of the actual importance of that term in the document) to give a measure of the importance of the term t_i within the particular document d_j . Thus we have the term frequency, defined as follows:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (11)$$

where $n_{i,j}$ is the number of occurrences of the considered term (t_i) in document d_j , and the denominator is the sum of number of occurrences of all terms in document d_j , that is, the size of the document $|d_j|$.

A threshold parameter is used here and changes during evaluation.

In the main system, queries attempt to find semantic contents such as specific people, objects and events in a broadcast news collection. We define the following classes: Named person, Named object and General object.

Our retrieval system needs to go through the following steps to find relevant multimedia resources for content-based queries without any user feedback or manual query expansion.

4 Evaluation

In this section we present the results of the experimental evaluation that we have conducted on the semantic relationships extraction using real datasets. The objective is to evaluate the efficiency of the schema transformation. We have used several resources sets attempting to cover many domain variations, features and special cases. The algorithm has been implemented using PHP language on the top of the open source, native XML database.

4.1 Accuracy Evaluation

We present the performance of the schema transformation generation processes of previous sections, then we examine the efficiency of the said processes.

The experiments were conducted on a 2.53GHz Intel Core I5 machine with 4GB of RAM, running MS 8. All results are averaged over three runs.

The basic characteristics (e.g., number of elements, attributes, etc.) of the XML resources are not shown in the evaluation. In this experiment, we have chosen a context about international politics and United States politics mainly but not exclusively. We have also used various contexts, as sports, terrorism, or Internet privacy.

Annotation has an important role here, there is a consequence and a difference between, for example, the keyword “Angela Merkel” appearing or not in “keyword” tag, and/or in “subject” tag too. Besides, taking for example, the keyword Hillary Clinton which could exist in other forms like Hillary Rodham Clinton or even Hillary Diane Rodham Clinton, has a real impact on results. This depends on news agencies, but actually, we did not deal with this point.

Table 1 shows results by number of used resources. Even a small number of resources is used, we believe that the use of a huge database could not have an impact on the results or on the performances. The result of the variation of used thresholds is presented in table 2. We can observe that by increasing it, redundancy decreases but the total number found decreases too.

Finally, as expected, the XML Schema file size slightly affects the time consumed by this transformation, e.g. a single iteration for a result set size of 20 resources takes about 3 seconds.

Table 1. Results values by modifying the number of resources

Resources	Total	Erroneous	Valid	Not detected
5	5	2	3	2
20	8	2	6	3
30	11	3	8	5

Table 2. Results values using variable thresholds

	Total	Not found	Redundancy
Similarity threshold >1	8	3	1
Similarity threshold >1.5	7	3	0

Table 3. Sample of the results values by relationship types

Relationship	Detected	Erroneous	Not detected
Show	1		1
Appear In	2	1	1
Talk	0		
Talk About	1		
Speak	1	1	
Speak About	3		1

In table 3, we present the results by relationship types. We assume and believe that the contents are the primary determinants of these results.

In fact, there is a lot of data which are made by humans, so the same content could be written in many ways hence influencing the interpretation even if made by humans. Consequently, this can result in weak structures. The fact of omitting Named Entity or describing differently a situation could change retrieval results.

Also the fact of using resources with the same context is very important, because otherwise, we could have zero relationships. Finally, the rank of metadata in XML resources is also computed and has an effect on results too.

e.g.:

```
<genre qcode="genre:WarConflict"></genre>
<genre qcode="genre:Politics"></genre>
```

These two tags do not have the same impact if the creator of the metadata considers the importance of the rank of those genres or chooses to put the inverse.

We note that we have more relationships between the same type of resources than between different types even those between videos and texts. We have an interesting number of relations. We can note that the factor “type” is important.

Notice that the poor descriptions of images has impacts on the results because usually images are not well annotated, seeing their nature. In addition, images have no <title> tag. Besides, the image sometimes describes a general context and does not specify persons or known entities, e.g. a picture containing scenes of injury or dead bodies.

Furthermore, we would measure the Recall as the fraction of the relations that are relevant and successfully retrieved, and the Precision as the fraction of retrieved documents that are relevant to the result obtained according to our perception.

$$Precision = \frac{|Correct\ detected\ relations|}{|Total\ retrieved\ relations|} \tag{12}$$

Precision: a fraction of documents that are relevant among the entire retrieved document. Practically it gives accuracy of the results.

$$Recall = \frac{|Correct\ detected\ relations|}{|Total\ correct\ relations|} \quad (13)$$

Recall: a fraction of the documents that is retrieved and relevant among all relevant documents. Practically it gives coverage of the results.

Precision is more important than Recall in our case, because the irrelevant/wrong relationships has negative repercussions on the results in our retrieval system. Figure 2 shows that the overall precision of our system is 0.69 indeed.

Consider also that normally, the number of used resources do not affects results. Performance will be quite close to this limit even if we increase this number. In some cases, for example, in the absence of connexion between resources, we could get 0 relationships which does not imply irrelevance of the algorithm.

In the next section, we introduce a new measure more meaningful than Recall and Precision.

4.2 Accuracy Metrics

It is well known, that the policy of the user providing relevance feedback can have a strong impact on the evaluation results. Since the user's views differ, judging the correctness of the retrieved relationships is a challenging task besides the distinction between relevant combination of relationships which is related to different interpretations.

In fact, the correctness of a detected relation is not a bivalent value as it is based on the user's perception. A relation could be irrelevant or missed. In essence, for an evaluation, a missed relation is better than an irrelevant one because this latter could have repercussions on the research results.

We employ and invite three testers to evaluate how closely the results satisfied their intentions.

For every user, we compare its interpretation with the original one. To assess the correctness of the algorithm, the results were manually examined by domain experts, and for this reason, we introduced a special measure that is Alg_i .

This measure allows a user to rate the correctness of each relationship and penalize the irrelevant ones, depending on its own perception.

Let us consider the following:

i, j : indexes of documents and k : number of relationships where:

$$\begin{matrix} i, j \\ i \neq j \end{matrix} \in]0,1[^2 \quad k \in \{1 \rightarrow 6\}$$

M : number of users

I : {indexes of all documents}

S : relations between I & j and

$$Alg_i\{L_{ij}^k\}$$

For every Alg_i we associate a subset E_{ij} (indexes of relations between fixed i and j)

$$Alg_{x1}: \{L_{ij}^s\}$$

$$Alg(x1) = C1 = C1^* + C1^{**}$$

To compute $C1$, we fixe X^* and X^{**} which are penalties

To penalize (-) $\leftrightarrow X^*$

To penalize (+) $\leftrightarrow X^{**}$

Where: $X^{**} \ll X^*$

01. In case of irrelevant relationship
02. Initializing $C1^* \leftarrow 0$
03. For $i, j \in I \ / i \neq j$
04. For $k \in E_{ij}$
05. For $s \in X_{ij}$ and $s \neq k : C1^* \leftarrow C1^* + X^*$
06. Return $C1^*$
07. In case of missed relationship
08. Initializing $C1^{**} \leftarrow 0$
09. For $i, j \in I \ / i \neq j$
10. For $s \in X_{ij}$
11. For $k \in E_{ij}$ and $s \neq k : C1^{**} \leftarrow C1^{**} + X^{**}$
12. Return $C1^{**}$

$$0 \leq C1 \leq 6X^*$$

$$X^* \in]0, 1[$$

$$0 \leq G = \frac{C1 + C2 + \dots + CM}{M} \leq 6X^*$$

$$0 \leq G \leq 1$$

We simply have to set X as a scalar to get results. According to our metric, the more the number tends towards zero, the more this number is relevant. In our case, X is set to $1/3$. We perceive through figure 3, that all obtained results are close and good. This has an important impact on ambitious efforts to detect relationships with more efficiency.

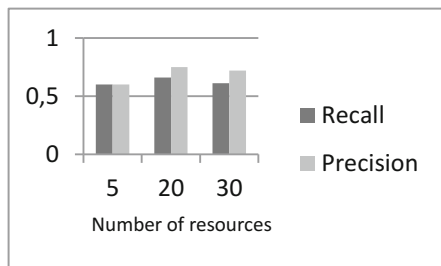


Fig. 2. Accuracy Recall & Precision

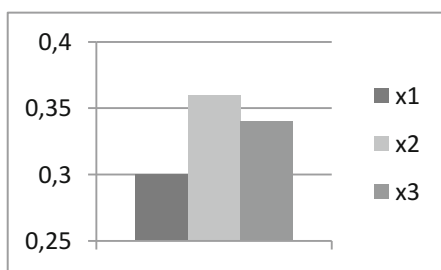


Fig. 3. Accuracy of Alg_i

5 Conclusion

Metadata provides rich semantic relationships that can be used for retrieval purposes. In order to capitalize hidden connexions and relationships between resources, we have presented in this paper a proposition for interlinking multimedia resources semantically through defined rules, and then, results are supplied as a contextual schema of RDF triples. The goal of this schema is primarily refining querying, and adding more semantics to our retrieval system.

The experimental results show that our approach can accurately find hidden relations between resources, and thus our main retrieval system will perform in a better way.

Actually, among obtained relations, there are some wrongly detected relations and some correct ones which are not detected. The next step is to continue exploring ways to improve the Precision of the construction of the relationships with poorer performance. In particular, the use of inference that may express ambiguous relationships, depending on the context, needs to be further enhanced. We also plan to try this algorithm with yet newer relationships based on resulting resources, we could build new relationships that will be used in second time.

We will continue investigating on the best combination of annotation and recommendation for using keywords to get better result.

References

1. Aouadi, H., Torjmen, M.: Exploitation des liens pour la recherche d'images dans des documents XML. In: Conférence Francophone en Recherche d'Information et Applications – CORIA (2010)
2. Bu, F., Hao, Y., Zhu, X.: Semantic Relationship Discovery with Wikipedia Structure. In: 22nd International Joint Conference on Artificial Intelligence IJCAI 2011, Barcelona, pp. 1770–1777 (2011)
3. Celik, I., Abel, F., Houben, G.-J.: Learning Semantic Relationships between Entities in Twitter. In: Auer, S., Díaz, O., Papadopoulos, G.A. (eds.) ICWE 2011. LNCS, vol. 6757, pp. 167–181. Springer, Heidelberg (2011)
4. Heim, P., Lohmann, S., Stegemann, T.: Interactive Relationship Discovery via the Semantic Web. In: Aroyo, L., Antoniou, G., Hyvönen, E., ten Teije, A., Stuckenschmidt, H., Cabral, L., Tudorache, T. (eds.) ESWC 2010, Part I. LNCS, vol. 6088, pp. 303–317. Springer, Heidelberg (2010)
5. Kharrat, M., Jedidi, A., Gargouri, F.: A system proposal for multimodal retrieval of multimedia documents. In: 9th IEEE International Symposium on Parallel and Distributed Processing with Applications. ISPA, Busan-Korea (2011)
6. Stanchev, L.: Building Semantic Corpus from WordNet. In: IEEE International Conference on Bioinformatics and Biomedicine Workshops, Philadelphia (2012)
7. Xu, M., Wang, Z., Bie, R., Li, J., Zheng, C., Ke, W., Zhou, M.: Discovering Missing Semantic Relations between Entities. In: Alani, H., et al. (eds.) ISWC 2013, Part I. LNCS, vol. 8218, pp. 673–686. Springer, Heidelberg (2013)