# Evaluation of Similarity Measures for Ontology Mapping

Ryutaro Ichise

National Institute of Informatics,
Tokyo 101-8430, Japan
ichise@nii.ac.jp

**Abstract.** This paper presents an analysis of similarity measures for identifying ontology mapping. Using discriminant analysis, we investigated forty-eight similarity measures such as string matching and knowledge based similarities that have been used in previous systems. As a result, we extracted twenty-two effective similarity measures for identifying ontology mapping out of forty-eight possible similarity measures. The extracted measures vary widely in the type in similarity.

## 1 Introduction

Many people now use the web to collect a wide range of information. For example, when making vacation plans, we check the web for lodging, routes, and sightseeing spots. Because these web sites are operated by individual enterprises, we have to search the sites manually to gather information. In order to solve such a problem, the Semantic Web is expected to become a next-generation web standard that can connect different data resources. On the Semantic Web, the semantics of the data are provided by ontologies for the interoperability of resources. However, since ontologies cover a particular domain or use, it is necessary to develop a method to map multiple ontologies for covering wide domains or different uses. Ichise organized an ontology mapping method for the interoperability of ontologies with a machine learning framework [1]. The framework uses a standard machine learning method with multiple concept similarity measures. Moreover, the paper defines many types of similarity measures, introduced from state-of-the-art systems. Although the system successfully integrates features for ontology mapping from different systems, we still do not know which features are effective for ontology mapping. In this paper, we present an experimental evaluation of a wide range of similarity measures in order to identify effective features for ontology mapping.

This paper is organized as follows. First, we discuss the problem of ontology mapping that we are undertaking and our approach using machine learning with multiple similarity measures. Next, we discuss the similarity measures for ontology mapping. Then, we evaluate the effectiveness of several similarity measures by using real data. Finally, we present our conclusions.

H. Hattori et al. (Eds.): JSAI 2008, LNAI 5447, pp. 15–25, 2009.

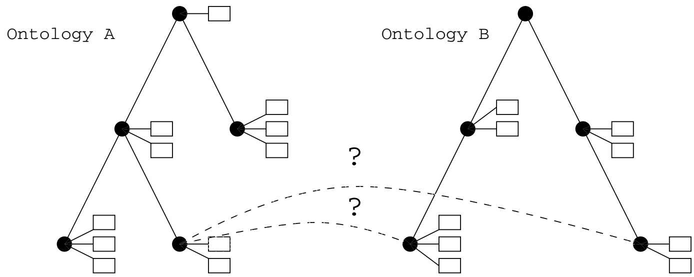


**Fig. 1.** Problem statement

## 2 Ontology Mapping

### 2.1 Problem

In this section, we describe the ontology mapping [2] that we are investigating. When we have many instances of objects or information, we usually use a concept hierarchy to classify them. Ontologies are used for such organization. We assume that the ontologies in this paper are designed for such use. The ontologies used for our paper can be defined as follows:

> The ontology $O$ contains a set of concepts, $C_1, C_2, \ldots, C_n$, that are organized into a hierarchy. Each concept is labeled by strings and can contain instances.

An example of an ontology is shown in the graph representation on the left side of Figure 1. The black circles represent a concept in the ontology and the white boxes represent instances in the ontology. The concepts (black circles) are organized into an hierarchy.

The ontology mapping problem can be stated as follows. When there are two different ontologies, how do we find the mapping of concepts between them? For example, in Figure 1, the problem is finding a concept in ontology B that corresponds to the concept in ontology A. For the concept at the bottom right side of ontology A, a possible mapping in ontology B can be the right bottom concept or the left bottom concept, or there may be others.

### 2.2 Machine Learning Approach for Ontology Mapping

In order to solve the ontology mapping problem, Ichise proposed to use the machine learning approach with multiple concept similarity measures [1]. In this section, we describe the method to convert the ontology mapping problem into a machine learning framework by using similarity measures.

To solve the ontology mapping problem, we think about the combination of concepts among different ontologies. In this case, the problem can be defining the value of a combination pair. In other words, the ontology mapping problem consists of defining the value of pairs of concepts in a concept pair matrix, as

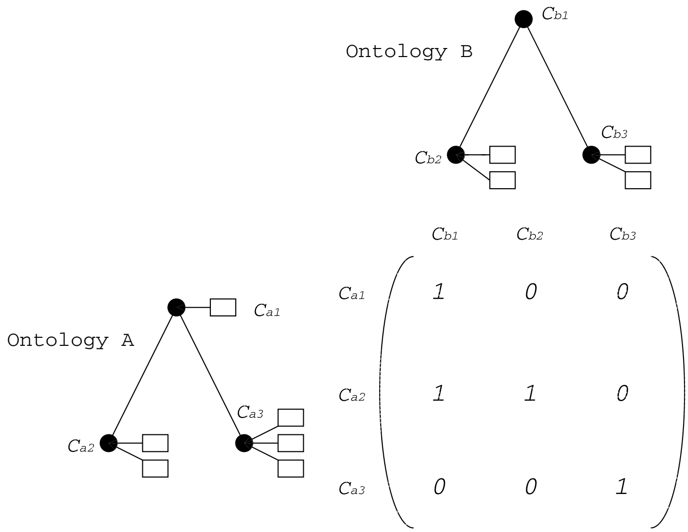


**Fig. 2.** Mapping matrix

shown in Figure 2. The rows of the matrix illustrate the concepts of Ontology A, that is, $C_{a1}, C_{a2}$ and $C_{a3}$, and the columns of the matrix illustrate the concept of Ontology B, that is, $C_{b1}, C_{b2}$ and $C_{b3}$. The values in the matrix represent the validity of the mapping. The value is 1 when two concepts can be mapped and 0 when two concepts cannot be mapped. For example, the second row and third column of the matrix represents the validity of mapping for $C_{a2}$ on Ontology A and $C_{b3}$ on Ontology B. This particular mapping isn't valid because the value in the matrix is 0.

The next question is what type of information is available to compose the matrix? According to our definition of ontologies, we can define a similarity measure of concepts by using a string matching method, such as concept name matching, and so on. However, a single similarity measure is not enough to determine the matrix because of the diversity of ontologies. For example, we can assume the concept of "bank" in two ontologies. The concepts seem to be mapped when we use the string similarity measure. However, when one ontology has a super concept of "finance" and another has that of "construction," these two concepts should not be mapped because each represents a different concept. In such a case, we should also use another similarity measure of the concepts.

From the above discussion, the problem is to define matrix values by using multiple similarity values of the concepts. As a result, we can write the problem in table form, as shown in Table 1. *ID* shown in the table represents a pair of concepts, *Class* represents the validity of the mapping, and the columns in the middle represent the similarity of the concept pairs. For example, the first line of the table represents the ontology mapping for $C_{a1}$ and $C_{b1}$ and has a similarity value of 0.75 for similarity measure 1. When we know some correct mappings, such as $C_{a1} \Leftrightarrow C_{b1}$ and $C_{a1} \Leftrightarrow C_{b2}$, we can use the mapping to determine

**Table 1.** Description of the ontology mapping problem

| ID | Similarity Measure 1 | Similarity Measure 2 | ... | Similarity Measure n | Class |
|---|---|---|---|---|---|
| $C_{a1} \Leftrightarrow C_{b1}$ | 0.75 | 0.4 | ... | 0.38 | 1 (Positive) |
| $C_{a1} \Leftrightarrow C_{b2}$ | 0.52 | 0.7 | ... | 0.42 | 0 (Negative) |
| ... | ... | ... | ... | ... | ... |
| $C_{a5} \Leftrightarrow C_{b7}$ | 0.38 | 0.6 | ... | 0.25 | ? |
| ... | ... | ... | ... | ... | ... |

the importance of the similarity measures. Then, we can make a decision on unknown classes such as $C_{a5} \Leftrightarrow C_{b7}$ by using the importance of the similarity measures. The example table shown represents the same problem that occurs in a supervised machine learning framework. Therefore, we can convert the ontology mapping problem into a normal machine learning framework.

## 3 Similarity Measures of Concepts

Many similarity measures have been proposed for measuring concept similarities. Examples include the string based similarity, graph based similarity, and knowledge based similarity. The string based similarity is widely used for ontology mapping systems. The graph based similarity utilizes the similarity of the structures of ontologies. In this measure, the ontologies are organized as tree structures, and so we can calculate the graph similarity of the ontologies. Examples include Similarity Flooding [3] and S-Match [4]. The knowledge based similarity utilizes other knowledge resources, such as a dictionary and WordNet [5], to calculate similarities.

Usually, ontology mapping systems utilize several types of similarity measures. For example, COMA++ [6] uses a matcher library, which corresponds to multiple similarity measures. However, most systems utilize similarity measures as static values. In other words, these systems do not weight the importance of similarity measures. In this paper, we investigate several of the similarity measures presented in [1] through experiments with real data. Our goal is to identify effective features. In the rest of this section, we discuss some definitions of similarity measures. We used four similarity measures. The similarities are "word similarity," "word list similarity," "concept hierarchy similarity," and "structure similarity." We will discuss these in this order.

### 3.1 Word Similarity

In order to calculate concept similarity, we introduce four string based similarities and also four knowledge based similarities as base measures.

The string based similarity is calculated for words. We utilize the following similarities:

– prefix
– suffix
– edit distance
– n-gram

The prefix similarity measures the similarity of word prefixes such as Eng. and England. The suffix similarity measures the similarity of word suffixes such as phone and telephone. Edit distance can calculate the similarity from the count of string substitution, deletion and addition. For n-gram, a word is divided into n number of strings, and the similarity is calculated by the number of same string sets. For example, similarity between "word" and "ward" is counted as follows. The first word "word" is divided into "wo, or, rd" for the 2-gram, and the second word "ward" is divided into "wa, ar, rd" for the 2-gram. As a result, we can find the similar string "rd" as the similarity measure for the 2-gram. In our system, we utilize the 3-gram for calculating similarity.

The knowledge based similarity is also calculated for words. We use WordNet as the knowledge resource for calculating similarity. Although a wide variety of similarities for WordNet are proposed, we utilize four:

– synset
– Wu & Palmer
– description
– Lin

The first similarity measure *synset* utilizes the path length of the synset in WordNet. WordNet is organized with synsets. Therefore, we can calculate the shortest path of different word pairs using synsets. The second similarity measure, Wu & Palmer, uses the depth and least common superconcept (LCS) of words [7]. The similarity is calculated in the following equation:

$$similarity(W_1, W_2) = \frac{2 \times depth(LCS)}{depth(W_1) + depth(W_2)}$$

$W_1$ and $W_2$ denote word labels for a concept pair, the depth is the depth from the root to the word and LCS is the least common superconcept of $W_1$ and $W_2$. The third similarity measure, *description*, utilizes the description of a concept in WordNet. The similarity is calculated as the square of the common word length in the descriptions of each word of a pair. The last similarity measure is proposed by Lin [8]. This measure is a calculation using a formula similar to that of Wu & Palmer except it uses information criteria instead of depth.

### 3.2 Word List Similarity

In this section, we extend the word similarity measures presented in the previous section. Word similarity measures are designed for words, but the measures are not applicable to a word list such as "Food_Wine." Such a word list can usually be used as a concept label. If we divide such words by a hyphen or underscore, we can obtain a word list. We define two types of similarities for a word list: *maximum word similarity* and *word edit distance*.

Let us first explain the maximum word similarity. When we use the combination of words in both lists, we can calculate the similarity for each pair of words by word similarity measures. We use the maximum value for word pairs in the word list as the maximum word similarity. In our paper, since we define eight word similarities (stated in the previous section), we can obtain eight maximum word similarities.

The second similarity measure, word edit distance, is derived from the edit distance. In the edit distance definition, similarity is calculated by each string. We extend this method, considering words as strings. Let us assume two word lists, "Pyramid" and "Pyramid, Theory." It is easy to see the two lists are very similar. If we consider one word as a component, we can calculate the edit distance for the word lists. In this case, "Pyramid" is the same in both word lists, so we can calculate the word edit distance as 1. Furthermore, if we assume "Top" and "Pyramid, Theory," the word edit distance is 2. As such, we can calculate the similarity by the word distance. However, another problem occurs for similar word lists. For example, when we assume "Social, Science" and "Social, Sci," the similarity is difficult to determine. The problem is the calculation of similarity for "Science" and "Sci, " that is, whether the two words are the same word. If we decide the two words are the same, the word edit distance is 0, but if not, the word edit distance is 1. In order to calculate the similarity of the words, we utilize the word similarity measure. For example, if we use the prefix as the word similarity measure, we can consider the two words as the same for calculating the word edit distance. However, if we use the synset as the word similarity measure, we cannot consider the two words as the same because "sci" is not in WordNet. From the above discussion, we can define the word edit distance for the eight word similarity measures. As a result, we define sixteen similarity measures for word lists, which include eight maximum word similarities and eight word edit distance similarities.

### 3.3 Concept Hierarchy Similarity

In this section we discuss the similarity for the concept hierarchy of an ontology. As we discussed in Section 2, ontologies are organized as concept hierarchies. In order to utilize the similarity of a concept hierarchy, we introduce similarity measures for concept hierarchies. The concept hierarchy similarity measure is calculated for the path from the root to the concept. Let us explain by the example shown in Table 2. We assume the calculation of the path "Top / Social_Sci" in ontology A and "Top / Social_Science" in ontology B. For calculation of the similarity, we divide the path into a list of concepts, as shown in the middle column of Table 2. Then, the similarity is calculated by the edit distance if we consider the concept as a component. For example, the concept "Top" is the same in both ontologies, but the second concept is different. Then, the edit distance for the path is 1. However, how do we decide whether the concept is the same or not? To determine this, we divide the concept into the word list for calculating the similarity by using the word list similarity. In this case, if "Social_Sci" and "Social_Science" are considered as a similar concept using the

**Table 2.** Examples for concept hierarchy similarity calculation.

| | Path | Path list | Word list |
|---|---|---|---|
| Ontology A | Top / Social_Sci | {Top, Social_Sci} | {Top}, {Social, Sci} |
| Ontology B | Top / Social_Science | {Top, Social_Science} | {Top}, {Social, Science} |

word list similarity, the edit distance is 0; if the two concepts are not considered as a similar concept using the word list similarity, the edit distance is 1. In other words, we calculate the edit distance with the right-hand lists in Table 2. As a result, we can calculate the concept hierarchy similarity by using the edit distance of the path. Because we can use any word list similarity measures for deciding the similarity of the word list, we obtain sixteen concept hierarchy similarity measures.

### 3.4 Structure Similarity

In this section, we define the similarity measures that use the structure of ontologies. In the previous section, we defined similarity using the concept hierarchy. However, a similarity can contain the similarity of a parent. We utilize the parent concept label for calculating similarity. This similarity is one of the variations of structure similarities, because it measures the neighborhoods on the graphs. Because the similarity is calculated by word list similarity, we can obtain 16 similarity measures for parents.

## 4 Evaluation of Similarity Measures

### 4.1 Internet Directory Data

In order to evaluate the effectiveness of similarity measures for the ontology mapping problem, we conducted an analysis of 48 similarity measures, which include 16 word list similarity measures, 16 concept hierarchy similarity measures, and 16 structure similarity measures. In our paper we used real Internet directory data, provided by the Ontology Alignment Evaluation Initiative (OAEI)[1] for the 2005 campaign. The data is constructed from three Internet directories, Google, Yahoo, and Looksmart, and contains simple relationships of class hierarchies. The data includes 2265 pairs of ontologies written in OWL format, and only one correct matching answer, which was verified by a human. Unfortunately, since the data has some format errors, we used 2193 pairs of ontologies and the correct mapping for the analysis. The data has positive (correct) mappings, and negative mappings are not available. We created negative mappings, as follows:

1. We choose the concept $C_s$, which is in the source ontology and has correct mappings.

[1] http://oaei.ontologymatching.org/

2. We randomly choose concept $C_t$, which is in the target ontology. If it is the correct mapping of $C_s$, then we again choose a concept in the target ontology.

As a result, the mapping pair produced by the above algorithm is relatively negative, not positive. We utilized the positive mappings and negative mappings for our experiments.

### 4.2 Analysis Method and Results

We conducted discriminant analysis for the 48 similarity measures to test the contribution of each similarity measure. In the analysis, we utilize the forward selection method, which takes, in order, the most effective explanatory variable into the discriminant. We utilize 5% as the level of significance when the variable is selected.

As a result, we obtained 22 similarity measures out of 48 possible similarity measures, as shown in Table 3. We consider those similarity measures as effective measures for identifying ontology mappings. On the left side of the table, *comparison target* denotes the type of objects compared in the ontologies. In this field, we have three values: "concept," "concept hierarchy," and "structure." The values come from comparison between concept labels defined in Section 3.2, comparison between concept hierarchies defined in Section 3.3, and comparison between structures defined in Section 3.4. *Word list method*, shown in the center of Table 3, denotes the type of word list similarity measure used. The value in this field has a "maximum word similarity" or "word edit distance," as defined in Section 3.2. *Base method*, shown on the right side of Table 3, denotes the base methods for comparing the similarity of words. There are eight possible values in this field: "prefix," "suffix," "edit distance," "n-gram," "synset," "Wu & Palmer," "description," and "Lin," as defined in Section 3.1.

When we examine the comparison target in Table 3, we can see the balanced results, which has 7 concepts, 8 concept hierarchies, and 7 structures. Although most previous systems for ontology mapping usually utilize the concept comparison, it is not enough to produce a good result. According to the results in Table 3, the other comparison targets, such as structure comparison and concept hierarchy comparison, are important for predicting the ontology mappings. However, when we examine the ranking of effective features, we see many features related to the comparison between concept hierarchies in the higher ranking of the list. Therefore, we can verify that the comparison between concept hierarchies is important for identifying a rough sketch of mappings, and the other comparisons are important for detail mappings.

Next, when we examine the word list method in Table 3, we see that the number of word edit distances is slightly larger than the number of maximum word similarities. The number of word edit distances is 13, and 9 for the maximum word similarity. However, the maximum word similarity appears in higher ranking of the list. Therefore, we can conclude that the maximum word similarity is effective for identifying rough mappings, but, the word edit distance is necessary to identify detail mappings.

**Table 3.** Effective similarity measures for ontology mapping

| Comparison target | Word list method | Base method |
|---|---|---|
| structure | maximum word similarity | edit distance |
| concept | maximum word similarity | edit distance |
| concept hierarchy | word edit distance | Lin |
| concept hierarchy | maximum word similarity | edit distance |
| concept hierarchy | word edit distance | description |
| concept hierarchy | maximum word similarity | description |
| concept hierarchy | word edit distance | prefix |
| concept hierarchy | maximum word similarity | Lin |
| concept hierarchy | maximum word similarity | synset |
| structure | maximum word similarity | Wu & Palmer |
| concept | word edit distance | n-gram |
| concept | maximum word similarity | Wu & Palmer |
| structure | word edit distance | Lin |
| concept hierarchy | word edit distance | Wu & Palmer |
| concept | word edit distance | Wu & Palmer |
| structure | word edit distance | description |
| structure | word edit distance | suffix |
| structure | word edit distance | synset |
| concept | maximum word similarity | description |
| concept | word edit distance | edit distance |
| concept | word edit distance | prefix |
| structure | word edit distance | prefix |

Finally, when we examine the base methods in Table 3, we can see balanced results: three prefixes, one suffix, four edit distances, one n-gram, two synsets, four Wu & Palmer's, four descriptions, and three Lin's. We can see from the results that all measures are necessary features for ontology mapping, because all methods appear in the list. In addition, we analyzed the measures for string-based similarity and knowledge-based similarity, which are discussed in Section 3.1. That list has 9 string based similarity measures and 13 knowledge based similarity measures. The number of knowledge based similarity measures are slightly larger then that of the string based similarity measures. The string based similarity measures are very popular in ontology mapping systems, but the knowledge based similarity measure is more effective for predicting the ontology mapping.

We define 48 similarity measures in this paper, but only 22 measures among them are effective for ontology mapping. The 22 measures consist of all the similarity measures defined in Section 3. The results indicate that there are no single definite method for identifying ontology mappings, and it is necessary to combine multiple methods for solving the ontology mapping problem.

We tested the obtained discriminant, which consists of 22 similarity measures. The results are shown in Figure 3. Group 1 and Group 2 denote positive examples (mapping) and negative examples (not mapping), respectively. The x-axis indicates the discriminant value and the y-axis indicates the percentage of examples. The accuracy is 73.78%. We understand that the problem is very difficult,

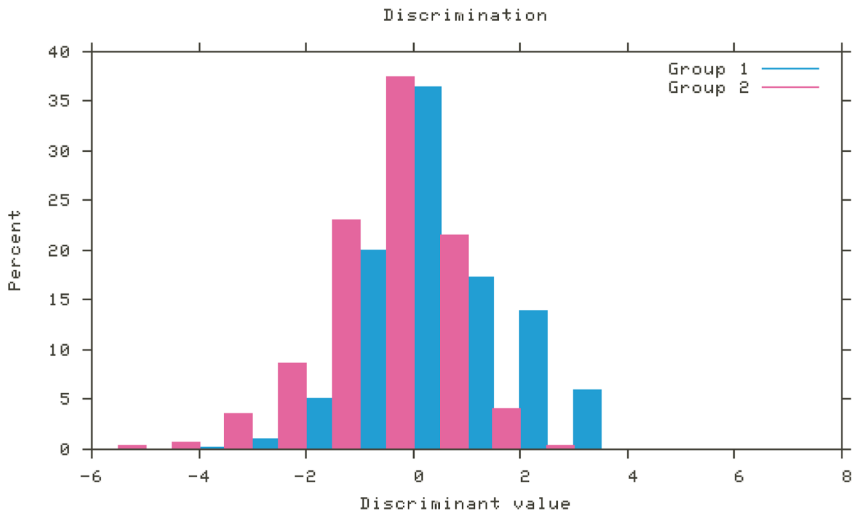


**Fig. 3.** Results of discrimination for ontology mapping

because the discrimination area in Figure 3 is very close between both groups. In addition, since the accuracy is not very high with the assumption of linear separation, we have to consider introducing a nonlinear learning method and more features, other than those discussed in this paper.

## 5 Conclusions

In this paper, we investigated effective features for deciding ontology mappings. We introduced several similarity measures and many types of similarities such as "word list similarity, " "concept hierarchy similarity" and "structure similarity." We analyzed these measures by discriminant analysis. As a result, we extracted 22 effective similarity measures out of 48 possible similarity measures. However, the extracted measures, which are effective for ontology mapping, vary widely in the type in similarity. Therefore, the experimental results suggest that for identifying ontology mapping it is necessary to use several types and compositions of similarity measures.

In our future work, we plan to extend the current research. In this work, we used 48 similarity measures introduced from previous research. However, there are still many other types of similarity measures for ontology mapping, such as instance based similarity measures [9]. In addition, Pedersen et al. proposed another measurement of word similarity [10]. We have to investigate such similarity measures, too. Furthermore, we would like to consider changing the learning methods. In this work, we utilized discriminant analysis for evaluating similarity measures. Since the performance was limited, the ontology mapping problem can be considered a nonlinear separation problem. Therefore, we would like to investigate several machine learning methods, including nonlinear machine learning methods, such as support vector machines (SVM) [11], to improve performance.

## References

1. Ichise, R.: Machine learning approach for ontology mapping using multiple concept similarity measures. In: Proceedings of the 7th IEEE/ACIS International Conference on Computer and Information Science, pp. 340–346 (2008)
2. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer, Heidelberg (2007)
3. Melnik, S., Garcia-Molina, H., Rahm, E.: Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In: Proceedings of the 18th International Conference on Data Engineering, San Jose, CA, pp. 117–128 (February 2002)
4. Giunchiglia, F., Shvaiko, P., Yatskevich, M.: S-match: an algorithm and an implementation of semantic matching. In: Bussler, C.J., Davies, J., Fensel, D., Studer, R. (eds.) ESWS 2004. LNCS, vol. 3053, pp. 61–75. Springer, Heidelberg (2004)
5. Fellbaum, C.: Wordnet: An Electronic Lexical Database. MIT Press, Cambridge (1998)
6. Aumueller, D., Do, H.H., Massmann, S., Rahm, E.: Schema and ontology matching with COMA++. In: Özcan, F. (ed.) Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 906–908. ACM, New York (2005)
7. Wu, Z., Palmer, M.: Verb semantics and lexical selection. In: Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, New Mexico State University, Las Cruces, New, Mexico, pp. 133–138 (1994)
8. Lin, D.: An information-theoretic definition of similarity. In: Proceedings of the 15th International Conference on Machine Learning, pp. 296–304. Morgan Kaufmann, San Francisco (1998)
9. Ichise, R., Takeda, H., Honiden, S.: Integrating multiple internet directories by instance-based learning. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence, pp. 22–28 (2003)
10. Pedersen, T., Patwardhan, S., Michelizzi, J.: Wordnet: similarity - measuring the relatedness of concepts. In: Proceedings of the 19th National Conference on Artificial Intelligence, pp. 1024–1025 (2004)
11. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, Cambridge (2000)