Wikification of Learning Objects Using Metadata as an Alternative Context for Disambiguation

Reyna Melara Abarca[1,2], Claudia Perez-Martinez[3], Alexander Gelbukh[1], Gabriel López Morteo[3], Magally Martinez Reyes[4], and Moisés Pérez López[4]

[1] Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico

[2] Escuela Superior de Cómputo (ESCOM), Instituto Politécnico Nacional, Mexico

[3] Instituto de Ingeniería, Universidad Autónoma de Baja California, Mexico

[4] Universidad Autónoma del Estado de México, Mexico

reynamelara@gmail.com, claudia.perez92@uabc.edu.mx, www.gelbukh.com

Abstract. We present a methodology to wikify learning objects. Our proposal is focused on two processes: word sense disambiguation and relevant phrase selection. The disambiguation process involves the use of the learning object's metadata as either additional or alternative context. This increases the probability of success when a learning object has a low quality context. The selection of relevant phrases is performed by identifying the highest values of semantic relatedness between the main subject of a learning object and the phrases. This criterion is useful for achieving the didactic objectives of the learning object.

Keywords. Word sense disambiguation, wikification, natural language processing, learning objects.

1 Introduction

Nowadays, educational resources such as Learning Objects (LO) include more graphical elements, which makes them more attractive to users by substantially reducing the textual content. Nevertheless, with this, some LO analysis processes turn out to be more difficult. The LO wikification involves word sense disambiguation (WSD), thus its precision is affected by the quality of the context.

In this paper, we introduce an approach for wikification of LO. The two most relevant aspects of this methodology include: (i) the way the WSD method is carried out with acceptable quality even without high quality context, and (ii) the fact that the method to select relevant phrases is based on semantic proximity of the phrases to the LO's main subject.

Considering that context quality is important to obtain assertive results in the WSD task [11, 16], the disambiguation process may lead to a non-assertive result when applied to a reduced textual content. We propose learning object wikification using a WSD method with a machine learning approach. Our methodology uses metadata as additional or alternative context.

In relation to selection of relevant phrases, our proposed method tries to accomplish educational goals by selecting phrases which are semantically closer to the main subject of the LO.

To estimate the usefulness of our proposal, we evaluated our WSD method on a corpus obtained by extracting approximately 60,000 training examples from Wikipedia. The accuracy of WSD was measured by k-fold cross-validation.

We observed two different cases in the results. In the first type of cases, when the context is sufficient, the accuracy of our algorithm has a slight improvement over the results of similar algorithms. This is due to the metadata acting as additional

ISSN 1405-5546
doi: 10.13053/CyS-18-4-2029

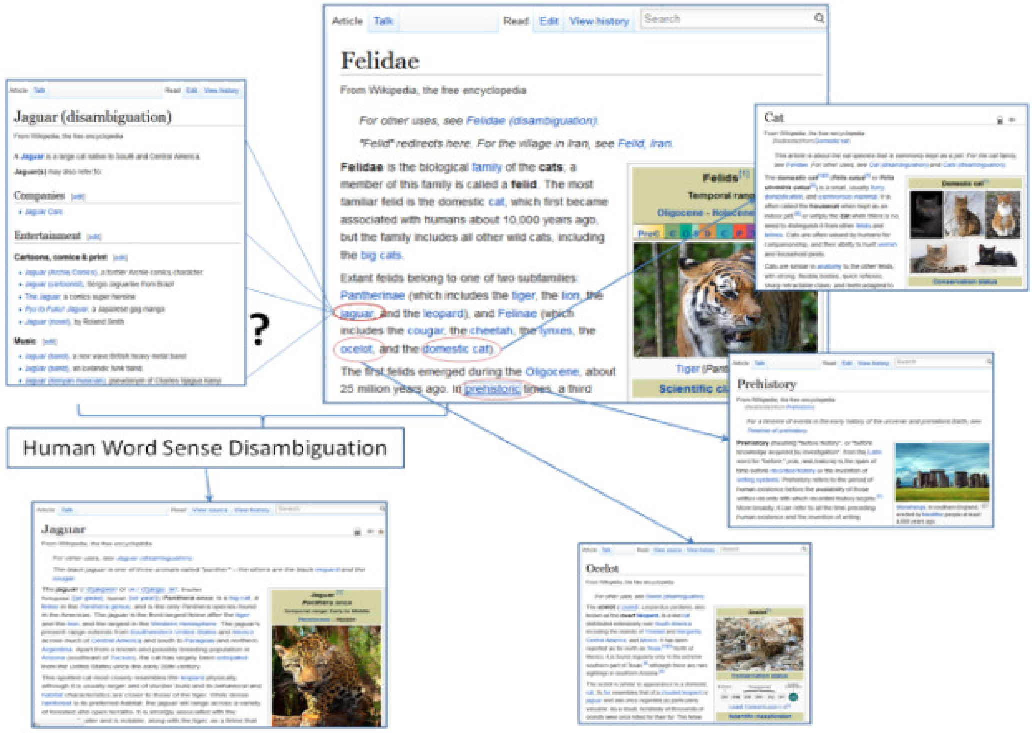

Fig. 1. Human WSD in wikification process

information. In the second case, when the context is not sufficient, its weakness is compensated by the metadata and, in contrast to other methods, the accuracy of our algorithm becomes very close to the case when the context is sufficient.

The proposed wikification process follows a task sequence which begins with extraction of useful information from the LO and finishes with delivering the wikified LO. Our approach has two main contributions. The first contribution is the disambiguation process based on the use of the metadata as either an alternative or additional context. This method slightly improves the result in the case of having sufficient quality context but it has a more relevant result when applied to poor-quality context. Our second contribution is the selection of the relevant phrases based on the measurement of their semantic relatedness with the LO's main subject, which contributes to the educational goals of the LO.

The paper is organized as follows. Section 2 describes related work on wikification, particularly, on LO wikification. Section 3 describes the proposed methodology. Section 4 contains the evaluation results for our WSD method. Finally, Section 5 presents conclusions and outlines some directions for future work.

2 Related Work

2.1 Wikification Process

The wikification process was inspired by *wikipedians,* people who edit the Wikipedia articles. They select relevant words or phrases in an article and link them to other Wikipedia articles whose titles correspond to these phrases. In certain cases there is more than one article that matches the phrase, so the appropriate article has to be selected according to the context. In this case, there is a disambiguation page with a list of possibilities.

As shown in Fig. 1, the phrase *jaguar* corresponds to more than one sense; therefore, in Wikipedia there is a disambiguation page, "Jaguar (disambiguation)", which lists several senses of the word *jaguar*. The wikipedians easily select the correct sense.

This process, which is so easy for humans, turns out to be very difficult if it is to be done automatically [1]. Text wikification is defined as a "*task of automatically extracting the most important words and phrases in the document, and identifying for each such keyword the appropriate link to a Wikipedia article*" [13]. The process involves two apparently easy tasks: selection of relevant phrases and WSD.

To decide which phrases are relevant, a wikipedian selects, according to his or her personal appreciation, the phrases that are important to explain. How can this task be automated in the form of an algorithm?

Concerning WSD, in a particular case of the wikification process, selecting the correct sense means selecting the adequate article from the disambiguation page for a phrase in a given context. Some works related to our approach are described below.

Mihalcea and Csomain [13] present a system that is able to automatically enrich plain text with links to encyclopedic knowledge. Two of the main processes involved in this are selection of relevant phrases to link and disambiguation of ambiguous phrases. In order to select relevant phrases, all possible n-gram candidates are extracted and weighted by the keyphraseness estimation method. To carry out the disambiguation process, the authors used the most frequent sense baseline WSD algorithm, a feature-based learning algorithm, and a combination of these two, obtaining the following results: a precision of 94.33%, a recall of 70.51%, and an F-measure of 80.69%.

To wikify, Milne and Witten [14] proposed a measure of semantic relatedness. This measure is obtained by using the hyperlink structure of Wikipedia as a source of knowledge. This measure is based on the Normalized Google Distance [6], which is in turn based on term occurrences on web pages. Specifically, their measure is specified as

$$relat(a,b) = \frac{\log(\max(|A|,|B|)) - \log(|A \cap B|)}{\log(|W|) - \log(\min(|A|,|B|))}, \qquad (1)$$

where A and B are the sets of all Wikipedia articles that link to a and b, respectively, W is the set of all Wikipedia articles, and $relat(a,b)$ is the semantic relatedness (better to say, distance) between the two Wikipedia articles a, b. Other interesting relatedness measures have also been proposed, including those that use vector-based or graph-based semantic representations [8, 10, 20, 23].

To carry out the wikification process, Milne and Witten [15] use machine learning for disambiguation. They visualize the Wikipedia links as a large set of manually defined connections between two manually disambiguated concepts. This structure can be used to learn to disambiguate phrases.

A Wikipedia article contains anchors or phrases that link to a target. The anchors correspond to either unambiguous or ambiguous phrases. The ambiguous phrases are found in their respective disambiguation pages within the Wikipedia database. While all possible candidate senses of ambiguous phrases can be found there, only one is manually selected as a target by the wikipedian. In Fig. 1, the anchor *jaguar* is an ambiguous phrase.

The method of WSD by means of machine learning proposed by Milne and Witten [15] shows how to build a training set to disambiguate links in a Wikipedia article. One example is created for each candidate sense. A positive example is the candidate sense selected by the wikipedian; the rest are negative examples.

Milne and Witten [15] built a three-featured vector for each example. The vector includes $Commoness\ (cs_i)$, or prior probability, which is the probability of that a candidate sense (cs_i) is used as a target in Wikipedia; $relatedness(cs_i, context)$, which is the semantic relatedness between a candidate sense and the context; and $quality\ contexty(cs_i)$, which is the sum of the semantic relatedness between a candidate sense and each element of the context; see more details in Section 3.2.

The context was formed by all links surrounding an ambiguous phrase within the given Wikipedia article. The performance of this approach was compared with previous works when using different classifiers. Their best results were a precision of 96.5%, a recall of 97.3%, and an F-

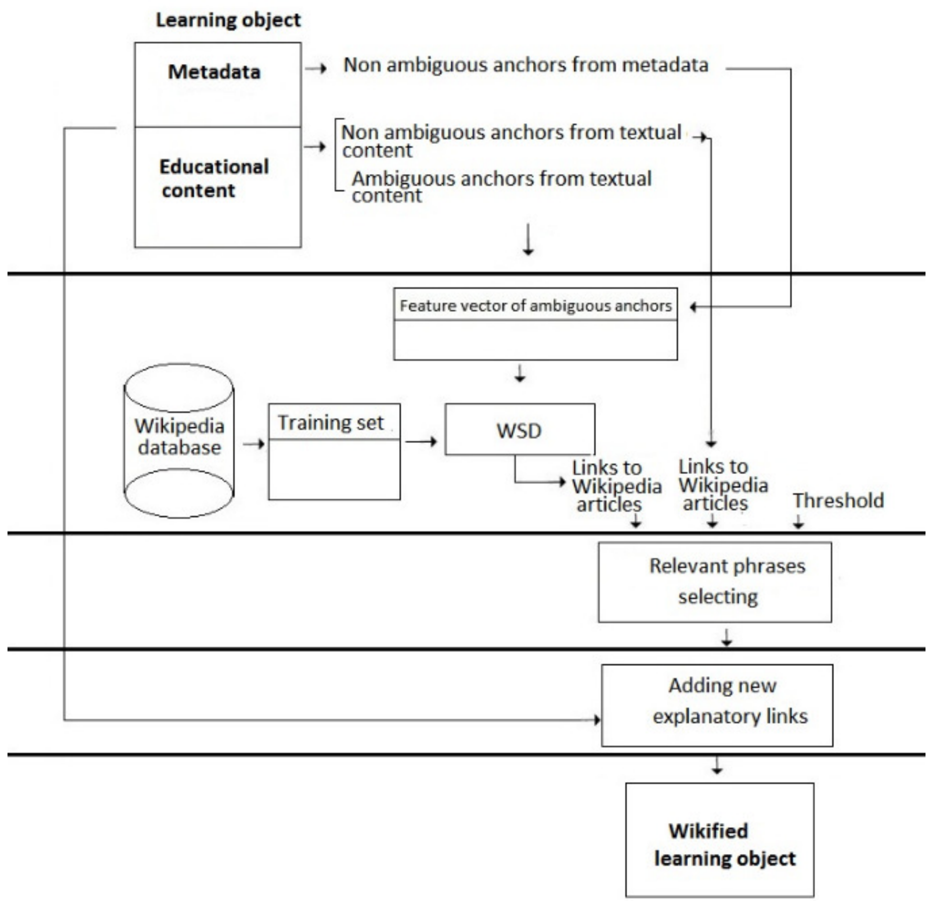

Fig. 2. Wikification process

measure of 96.9% when the C4.5 bagging algorithm was applied.

In our proposal, we used a modified version of the WSD approach proposed by Milne and Witten to disambiguate phrases. We added other criteria to the training set and obtained higher accuracy, as described in Section 4.

Cai et al. [2] carried out the wikification process according to a different approach, enriching articles by adding additional links to them. Instead of word sense disambiguation, they performed phrase sense disambiguation by applying the link co-occurrence matrix approach. Their sample included about 10,000 popular Wikipedia articles processed with moderate computational resources. They achieved a precision of 89.97% and a recall of 76.43%.

2.2 Wikification of Learning Objects

For our work, we adopted a more descriptive definition of a learning object than the one given in [22]: an LO is “a digital self-contained and reusable entity with a clear educational purpose, with at least three internal and editable components: content, learning activities and elements of context. The learning objects must have an external structure of information to facilitate their identification, storage and retrieval: the metadata” [5].

The metadata is an XML file containing descriptors or information about a Learning Object. It usually follows the IEEE LOM standard [12], but there are other important metadata standards such as SCORM and Dublin Core. All metadata standards specify in their structure the following two descriptors: title and description. Under this view of LO, some related works on wikification are described below.

Coursey et al. [7] suggested metadata generation by learning object repositories. In their paper, several methods for automatic keyword extraction are presented; the results of their evaluation demonstrate that automatic keyword extraction is a viable solution for suggesting terms and phrases for metadata annotation. They presented and tested different methods for keyword extraction, such as TextRank, wikifier, intersection, union, and Longest Common Substring (LCS).

3 Wikification of Learning Objects

The wikification process presented in this paper follows a sequence of tasks that begins with extraction of useful information from LO (metadata and textual content) and ends with delivering the LO with explanatory links toward Wikipedia articles; see Fig. 2.

This approach provides two main contributions: first, a WSD method based on the use of metadata as either an additional or alternative context, and second, a method to discriminate relevant phrases based on the degree of semantic relatedness with the LO's main subject. These two methods are described below.

3.1 Pre-processing

At the pre-processing stage, useful information is extracted from the LO's metadata and the educational content. The LO metadata contain a set of descriptors, but we select only the title and description. The metadata and the textual content of a learning object are parsed to look for phrases that match Wikipedia articles; each phrase can be either ambiguous or non-ambiguous.

Four sets of phrases are built: $MetadataDescrip$, $MetadataTitle$, $AmbiguousContent$, and $NonAmbiguousContent$. Only non-ambiguous phrases are selected from the metadata. The four sets are described as follows:

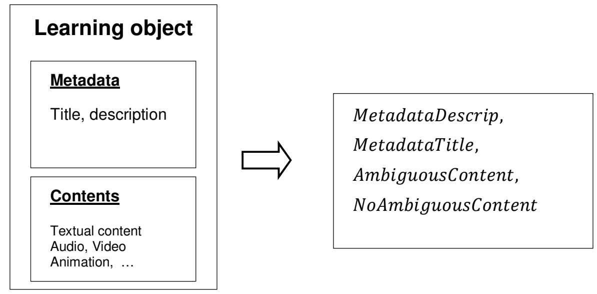

Fig. 3. Extracting useful information for disambiguation

$$MetadataDescrip = \{linkd_i | d_i \ match\ with\ phrase\ \in LO\ description\}, \quad (2)$$

$$MetadataTitle = \\ = \{linkd_i | d_i \ match\ with\ phrase\ \in\ LO\ title\}, \quad (3)$$

$$AmbiguousContent = \{link\ A_i | A_i match\ with\ textual\_\ content\ \&\ A_i\ is\ ambiguous\ phrase\}, \quad (4)$$

$$NonAmbiguousContent = \{link\ NA_i | NA_i\ match\ with\ textual\_content\ \&\ NA_i\ is\ non\ ambiguous\ phrase\}. \quad (5)$$

Ambiguous phrases are those that belong to a disambiguation page in the Wikipedia database; this means that there is more than one article that has the referred phrase in its title. This set of ambiguous phrases is disambiguated in the next phase. The sets of metadata and non-ambiguous content are useful to build a featured vector of the set of ambiguous content; it is disambiguated by the classifier as described below.

3.2 Word Sense Disambiguation (WSD)

Word sense disambiguation is the ability to computationally determine which sense of a word is activated by its use in a particular context [11]. To wikify an LO, a WSD process is necessary. WSD is carried out by extracting relevant phrases from the LO text and linking them toward the appropriate article, this involves selection among various candidate articles.

WSD in Wikipedia articles is easily carried out by humans, some phrases are not ambiguous and linked to Wikipedia articles that match them, for example, *prehistory*, *big cat,* and *ocelot*; however,

ISSN 1405-5546
doi: 10.13053/CyS-18-4-2029

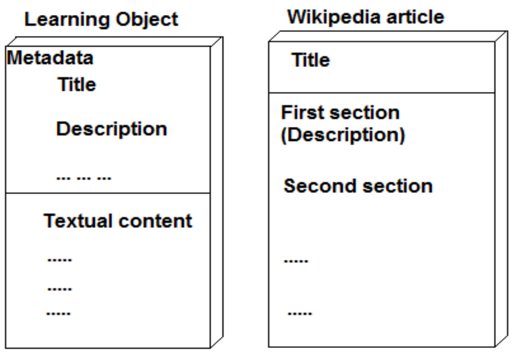

Fig. 4. Matching structures of LO and an article

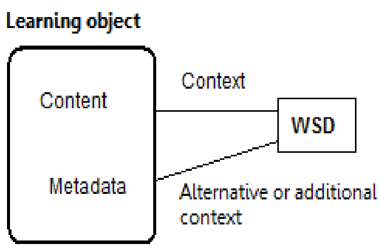

Fig. 5. WSD using alternative context

jaguar is an ambiguous phrase because it has more than one associated article (candidate sense) stored in the disambiguation page *Jaguar (Disambiguation)*, so a human wikipedian selects the correct sense according to the context of the phrase (human WSD).

For each edited article, these assertive selections of the sense are stored in the database of Wikipedia. Milne and Witten [15] proposed to extract this information by applying machine learning and a training set using three features (prior probability, relatedness between a candidate sense and a given ambiguous phrase, and quality context), with the C4.5 algorithm as a classification method.

Our WSD approach introduces changes to this algorithm: we add three features and use a different classifier algorithm based on the considerations that follow.

First, a Wikipedia article has a well-defined structure: it has the first section, which is described—according to the Manual of Style of Wikipedia[1]—as a section that "... *serves as an introduction to the article and a summary of its most important aspects... as a concise overview.*"

Therefore, the first section can be semantically closer to the correct candidate sense of any ambiguous phrase in the text. The Introduction section in a Wikipedia article matches the description of LO, so we can use it as another source of information for disambiguation purposes.

We propose to add an additional feature to be used for the classification task: the relatedness between a candidate sense for an ambiguous phrase and the title of the LO that contains it.

Second, another relevant piece of data is the title of LO which presents its content. For a Wikipedia article in the example given in Fig. 1, it would be possible to disambiguate *jaguar* only with the help of the title *felidae*, but this is not always the case. In any case, a title can complement the context and assist in disambiguation. The titles of the Wikipedia articles can also match the title in the LO structure, therefore, we can use them as features. So we add the feature relatedness, relatedness between a candidate sense for one ambiguous phrase and the description section. We can view the metadata either as an alternative context or an additional context to the WSD process.

Third, we add another feature, the average of all the attributes; this seems to reinforce the classification process and improves its accuracy as it can be seen in the section that presents evaluation.

Finally, concerning the method to compute the semantic relatedness, as it can be seen in (1), it behaves as a geometrical distance between two words: when they are semantically close, the relatedness (in fact, distance) measure is small. So we applied the nearest neighbor algorithm, in addition to other algorithms that previous authors have used.

The training set for the WSD process is presented in Fig. 6.

We calculate each feature according to the formulas:

[1]See http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Lead_section#cite_note-1

ISSN 1405-5546
doi: 10.13053/CyS-18-4-2029

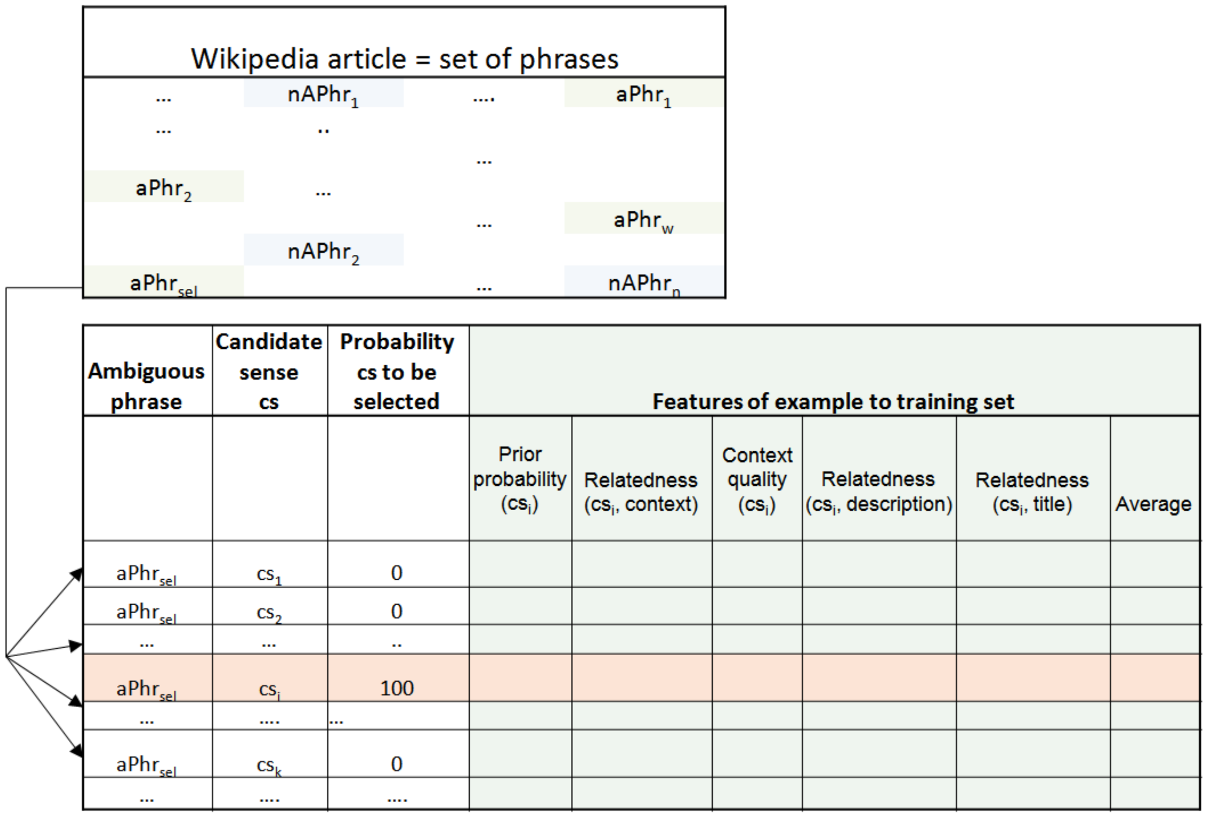

Fig. 6. Training set supplied to classifier for WSD

$$context = \{phrase\ NA_1, \ldots, phrase\ NA_m\}, \quad (6)$$

$$contextDescrip = \{phrase\ NAD_1, \ldots, phrase\ NAD_r\}, \quad (7)$$

$$prior\Pr obability(cs_i) = \frac{\#articles\ that\ use\ \ cs_i\ \ as\ anchor}{\sum \#articles\ that\ contain\ cs_i}, \quad (8)$$

$$relat_{(CS_i,\ context)} = \frac{\sum_{i=1}^{m} relatedness(CS_i, phraseNA_j)}{m}, \quad (9)$$

$$contextquality(cs_i) = \sum_{i=1}^{k} relatedness(cs_i, context). \quad (10)$$

In our method, three features are added to the vector:

$$relat_{(CS_i,\ title)},\ calculated\ by\ (1), \quad (11)$$

$$relat_{(CS_i,\ descrip)} = \frac{\sum_{r=1}^{t} relatedness(CS_i, phraseNAD_r)}{r}, \quad (12)$$

$$average\ (cs_i) = average\ of\ five\ features\ for\ cs_i. \quad (13)$$

The first three features are the same that were proposed in [15], as mentioned in Section 2.1. The two new features (semantic relatedness between a candidate sense and the title, and the relatedness between a candidate sense and the description) become very useful when the context quality is not sufficient for WSD.

Once the classifier has been trained, the ambiguous phrases from LO can be disambiguated. For this purpose it is necessary to represent each element of the $AmbiguousContent$ set with a feature vector in order to introduce them into the WSD module.

3.3 Relevant Phrases Selection

The relevant phrases selection process is based on an LO educational purpose targeted to reading

ISSN 1405-5546
doi: 10.13053/CyS-18-4-2029

Table 1. Training set for WSD

Sample		Prior probability	Relatedness (csi, description)	Relatedness (cs1,context)	Relatedness (csi, title)	Quality of context	Quality of description	Prob. to be selected	Accuracy
1	M&W	**		*		*		c	98.97
2	Our proposal	**	*	*	*	*	*	c	99.01
3	M&W without context	**						c	97.39
4	Our proposal without context	**	*		*			c	97.57

comprehension. Guthrie refers to the Concept-Oriented Reading Instruction as an instructional framework that supports cognitive strategies for knowledge construction during reading [9].

The objective is that a student would fully and deeply understand the text content if she is led to consciously add new information to her initial knowledge base by expanding the conceptual structures. The theoretical model states that instructional context enhances student engagement in reading, and reading engagement in its turn enhances reading achievement.

When a student finishes a reading exercise, she must have identified **core concepts** to which materials are directed, that is why this approach is called Concept-Oriented Reading Instruction.

These core concepts can be selected by wikification, and they can be closer to the main subject of LO. We propose to select all possible links, namely, all the phrases (anchors) that match a Wikipedia article, without any discrimination. Each anchor can be targeted to a Wikipedia article by one link, but first we must list and order them according to semantic relatedness between each of them and the main subject of the LO, that is, the title.

The threshold of acceptance can be selected by the user according to her purpose, for example, the purpose may be to carry out a reading comprehension strategy.

3.4 Adding New Explanatory Links

Once the possible links have been disambiguated and the selection based on the relatedness with the main subject has been done, what is left is to add the links to the LO structure; this process must be carried out according to the particular structure of the learning object.

4 Evaluation

We evaluated our method by accuracy of the WSD process. The training set was built by means of information extraction from the Wikipedia database using API JWPL [24]. As in previous works, we used approximately 60,000 examples extracted from 600 random Wikipedia articles, the Wikipedia version as of November 2013 was used.

As explained previously, we proposed two new features to resolve disambiguation based on machine learning; see Table 1. As it can be seen, the last attribute is used as the class; the classifier delivers two possible values: 1 if a given example is the adequate candidate sense in the context, otherwise the value is 0.

We found some interesting results; see Fig. 7. The first result of WSD using the approach in [15] is 98.97, if there is no context the results are less accurate. A slightly better result is obtained by the new approach in Sample 2. Also, a better result is

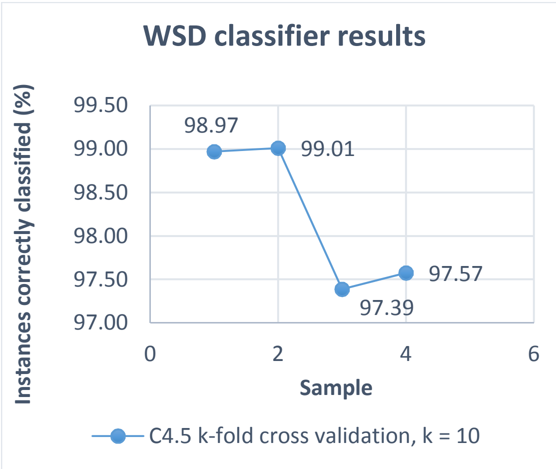

Fig. 7. Evaluating the disambiguation process

obtained in normal conditions when a good quality context is not present.

Therefore, this approach can be improved if more metadata is included.

5 Conclusions

We have presented a new wikification learning object methodology, which includes the use of metadata as either an additional or alternative context. Applying machine learning, we have tested various classifier algorithms; the best results were obtained by C4.5 algorithm using cross validation method.

The use of educational resources as an object of study within the Natural Language Processing task is a promising field, which can contribute to educational virtual environments.

In our future work, we plan to experiment with new relatedness measures and text representations. In particular, a promising direction to explore is linguistic-based n-gram representation of text [3, 17, 18, 19].

Acknowledgements

This work was done under partial support of the Mexican Government (CONACYT, SNI, COFAA-IPN, SIP-IPN 20144534) and FP7-PEOPLE-2010-IRSES: "Web Information Quality – Evaluation Initiative (WIQ-EI)" European Commission project 269180.

References

1. **Alonso-Rorís, V.M., Santos Gago, J.M., Pérez Rodríguez, R., Rivas Costa, C., Gómez Carballa, M.A., & Anido Rifón, L. (2014).** Information Extraction in Semantic, Highly-Structured, and Semi-Structured Web Sources. *Polibits*, Vol. 49, pp. 69–75.
2. **Cai, Z., Zhao, K., Zhu, K., & Wang, H. (2013)**. Wikification via link co-occurrence. *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management* (CIKM '13), ACM, New York, pp. 1087–1096.
3. **Calvo, H., Segura-Olivares, A., & García, A. (2014).** Dependency vs. Constituent Based Syntactic N-Grams in Text Similarity Measures for Paraphrase Recognition. *Computación y Sistemas,* Vol. 18, No. 3, pp. 517–554.
4. **Chiappe, A. (2009).** Acerca de lo pedagógico en los objetos de aprendizaje-reflexiones conceptuales hacia la construcción de su estructura teórica. *Estudios Pedagógicos,* Vol. XXXV, No. 1, pp. 261–272.
5. **Chiappe, A., Segovia, Y., & Rincon, H.Y. (2007).** Toward an instructional design model based on learning objects. *Educational Technology Research and Development,* Vol. 55, pp. 671–681.
6. **Cilibrasi, R.L. & Vitanyi, P.M.B. (2007).** The Google Similarity Distance. *IEEE Trans. Knowledge and Data Engineering*, Vol. 19, No. 3, pp. 370–383
7. **Coursey, K., Mihalcea, R, & Moen, W. (2008).** Automatic Keyword Extraction for Learning Object Repositories. *Proceedings of the American Society for Information Science and Technology*, Vol. 45, No. 1, pp. 1–10.
8. **Das, N., Ghosh, S., Gonçalves, T., & Quaresma, P. (2014).** Comparison of Different Graph Distance Metrics for Semantic Text Based Classification. *Polibits*, Vol. 49, pp. 51–57.
9. **Guthrie, J., Wigfield, A., Barbosa, P., Perencevich, K., Taboada, A., Davis, M., Scafiddi, N., & Tonks, S. (2004)**. Increasing Reading Comprehension and Engagement through Concept-Oriented Reading Instruction. *Journal of Educational Psychology,* Vol. 96, No. 3, pp. 403–423.

ISSN 1405-5546
doi: 10.13053/CyS-18-4-2029

10. **Huynh, D., Tran, D., Ma, W., & Sharma, D. (2014).** Semantic Similarity Measure Using Relational and Latent Topic Features. *International Journal of Computational Linguistics and Applications,* Vol. 5, No. 1, pp. 11–26.
11. **Indurkhya, N. & Damerau, F. (eds) (2010),** *Handbook of Natural Language Processing.* 2nd edition. CRC Press, New York.
12. **Koutsomitropoulos, D.A., Alexopoulos, A., Solomou, G., & Papatheodorou, T. (2010).** The Use of Metadata for Educational Resources. *Digital Repositories: Practices and Perspectives. D-Lib Magazine,* Vol. 16.
13. **Mihalcea, R. & Csomain, A. (2007).** Wikify!: linking documents to encyclopedic knowledge. *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM '07)*, ACM, New York, pp. 233–242.
14. **Milne, D. & Witten, I. (2008)**. *An Effective, Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links*. Association for the Advancement of Artificial Intelligence.
15. **Milne, D. & Witten, I. (2008).** Learning to link with Wikipedia. *Proceedings of the 17th ACM conference on Information and knowledge management (CIKM '08),* ACM, New York, NY, USA, pp. 509–518.
16. **Navigli, R. (2009).** Word Sense Disambiguation: A survey. *ACM Comput. Surv*., Vol. 41, No. 2, Article 10.
17. **Sidorov, G. (2013).** N-gramas sintácticos no-continuos. *Polibits,* Vol. 48, pp. 69–78.
18. **Sidorov, G. (2013).** Syntactic Dependency Based N-grams in Rule Based Automatic English as Second Language Grammar Correction. *International Journal of Computational Linguistics and Applications,* Vol. 4, No. 2, pp. 169–188.
19. **Sidorov, G. (2014).** Should Syntactic N-grams Contain Names of Syntactic Relations? *International Journal of Computational Linguistics and Applications,* Vol. 5, No. 1, pp. 139–158.
20. **Sidorov, G., Gelbukh, A., Gómez-Adorno, H., & Pinto, D. (2014).** Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model. *Computación y Sistemas,* Vol. 18, No. 3, pp. 491–504.
21. **Wigfield, A. (2004).** *Motivating Reading Comprehension: Concept-Oriented Reading Instruction.*
22. **Wiley, D.A. (2000).** Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D.A. Wiley (ed.), *The Instructional Use of Learning Objects.* [online].
23. **Wolf, L., Hanani, Y., Bar, K., & Dershowitz, N. (2014).** Joint word2vec Networks for Bilingual Semantic Representations. *International Journal of Computational Linguistics and Applications*, Vol. 5, No. 1, pp. 27–44.
24. **Zesch, T., Müller, C., & Gurevych, I. (2008).** Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. *Proceedings of the 6th International Conference on Language Resources and Evaluation.*

Reyna Melara Abarca is a professor of Superior School of Computing (ESCOM) of the National Polytechnic Institute (Instituto Politécnico Nacional, IPN), Mexico, and a PhD student at the Centro de Investigación en Computación, IPN.

Claudia Perez-Martinez received her Master degree in Computer Science from the Mexico State Autonomous University (Universidad Autónoma de Estado de México). She received her Master degree in Educational Technology from the Monterrey Institute of Technology (Instituto Tecnológico de Monterrey). She is pursuing a Ph.D. in Sciences at the Institute of Engineering, Baja California Autonomous University (Instituto de Ingeniería, Universidad Autónoma de Baja California).

Alexander Gelbukh received an MSc degree in Mathematics from the Lomonosov Moscow State University, Russia, and a Ph.D. in Computer Science from VINITI, Russia. He is currently a Research Professor and Head of the Natural Language Processing Laboratory of the Center for Computing Research (Centro de Investigación en Computación, CIC) of the National Polytechnic Institute (Instituto Politécnico Nacional, IPN), Mexico, and is the President of the Mexican Society of Artificial Intelligence (SMIA). He is a Member of the Mexican Academy of Sciences and National Researcher of Mexico (SNI) at Excellence level 2. He is author, coauthor, or editor of more than 500 research publications in natural language processing and artificial intelligence.

Gabriel López Morteo received a PhD degree in Computer Science from the Ensenada Center of

Scientific Research and Higher Education (Centro de Investigación Científica y de Educación Superior de Ensenada). He has academic experience in computer assisted education, collaborative systems, process engineering, and Internet application development. He is full time researcher at Department of Computing and Informatics of Institute of Engineering of Baja California Autonomous University (Departamento de Computación e Informática del Instituto de Ingeniería de la Universidad Autónoma de Baja California). He carries out research in the field of computer science and applied informatics oriented to electronic learning environments, interactive learning objects, and the behavior of learning virtual communities. He is coordinator of the Physics Engineering Area of the Baja California Institute of Engineering in Information Technologies (Instituto de Ingeniería en Tecnologías de la Información de Baja California, A.C.).

Magally Martinez Reyes received a PhD degree in Mathematics Education from Centro de Investigación y Estudios Avanzados of Instituto Politécnico Nacional. She has a Bachelor degree and a Master degree in Mathematics by Universidad Nacional Autónoma de México. She is a member of Sistema Nacional de Investigadores. Her research interest is development of educational software for learning mathematics. She is currently a full professor with PROMEP profile of the Universidad Autónoma del Estado de México Valle de Chalco, associated to the program of Computer engineering and MS in Computer Science. She published several specialized articles in national and international journals on mathematics and computing.

Moises Pérez López is an intern pursuing a bachelor degree in Computer Engineering from Universidad Autónoma del Estado de México Valle de Chalco. He is collaborating in the research project “Collaborative Spaces for research using Wikipedia”.

Article received on 27/08/2014; accepted on 27/11/2014.