Building Specialized Multilingual Lexical Graphs Using Community Resources

Mohammad Daoud, Christian Boitet, Kyo Kageura, Asanobu Kitamoto, Mathieu Mangeot, Daoud Daoud

Building Specialized Multilingual Lexical Graphs Using Community Resources

Mohammad Daoud [1], Christian Boitet [1], Kyo Kageura [2], Asanobu Kitamoto [3], Mathieu Mangeot [1], Daoud Daoud [4]

[1] Grenoble Informatics Laboratory, GETALP, Université Joseph Fourier, 385, rue de la Bibliothèque, 38041 Grenoble, France. {Mohammad.Daoud , Christian.Boitet, Mathieu.Mangeot}@imag.fr
[2] Library and Information Science Laboratory, Graduate School of Education, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan. kyo@p.u-tokyo.ac.jp
[3] The National Institute of Informatics. 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430. Kitamoto@nii.ac.jp
[4] Princess Sumaya University, P. O. Box 1438 Al-Jubaiha 11941 Jordan. Daoud@batelco.jo

Abstract. We are describing methods for compiling domain-dedicated multilingual terminological data from various resources. We focus on collecting data from online community users as a main source, therefore, our approach depends on acquiring contributions from volunteers (explicit approach), and it depends on analyzing users' behaviors to extract interesting patterns and facts (implicit approach). As a generic repository that can handle the collected multilingual terminological data, we are describing the concept of dedicated Multilingual Preterminological Graphs MPGs, and some automatic approaches for constructing them by analyzing the behavior of online community users. A Multilingual Preterminological Graph is a special lexical resource that contains massive amount of terms related to a special domain. We call it preterminological, because it is a raw material that can be used to build a standardized terminological repository. Building such a graph is difficult using traditional approaches, as it needs huge efforts by domain specialists and terminologists. In our approach, we build such a graph by analyzing the access log files of the website of the community, and by finding the important terms that have been used to search in that website, and their association with each other. We aim at making this graph as a seed repository so multilingual volunteers can contribute. We are experimenting this approach with the Digital Silk Road Project. We have used its access log files since its beginning in 2003, and obtained an initial graph of around 116000 terms. As an application, we used this graph to obtain a preterminological multilingual database that is serving a CLIR system for the DSR project.

1 Introduction

Discovering and translating domain specific terminology is a very complicated and expensive task, because (1) it depends on human terminologists [1], which increases the cost, (2) terminology is dynamic [2], thousands of terms are coined each year, and

(3) it is difficult to involve subject matter experts in the construction process. That will not only increase the cost, but it will reduce the quality, and the coverage (linguistic and informational) of the produced term base. Databases like [3-5] are built by huge organizations, and it is difficult for a new domain with a smaller community to produce its own multilingual terminological database.

There is some work on constructing lexical resources by using Machine-readable dictionaries "MDRs" [6] [7], however for a domain-specific terminology, the involvement of the community and domain experts is essential, and associating several multilingual repositories into a specialized database may affect the integrity of the data and the domain relevance. TransGraph [8] is another attempt to associate various MDRs into a graph of words and its translations, effective for finding translation equivalences for general purpose lexical units. However, such a graph cannot handle relations between terms for a specific domain. Besides, MDRs do not suffice to determine such relations between lexical units available in the terminological sphere of a domain.

We are trying to analyze various resources in order to replace the traditional way of extracting related terminology. We introduce the concept of multilingual preterminological graphs, which are constructed by analyzing the interaction between domain-related resources on one side, and domain experts on the other side. Basically, we analyze the access log files to find important terms used to access the website, and relations between them. This approach falls under the category of implicit user contribution. reCAPTCHA [9] is an example of using this kind of contribution. After constructing the initial graph, we try to multilingualize it by using online multilingual resources at the beginning, and then by accepting progressive enhancements from community users in an explicit contribution approach.

Multilingual knowledge in a specific domain may not be available in any format (MDRs, printed dictionaries…). But such knowledge might be known and used by specialized multilingual people. We claim that discovering them, and encouraging them to contribute (explicitly and implicitly) is as important as discovering digital resources, or web services…

The remainder of this paper is organized as follows. The second section introduces the MPGs and the implicit and explicit approaches to construct them. The third section describes the extraction and contribution platform and its applications. Then section four reports the experimental results. And finally, section five draws some conclusions.

2 Multilingual Preterminological Graphs: Construction and Evolution

2.1 Definitions

We begin by describing multilingual preterminological graphs in detail, and present the approaches to initialize and multilingualize them.

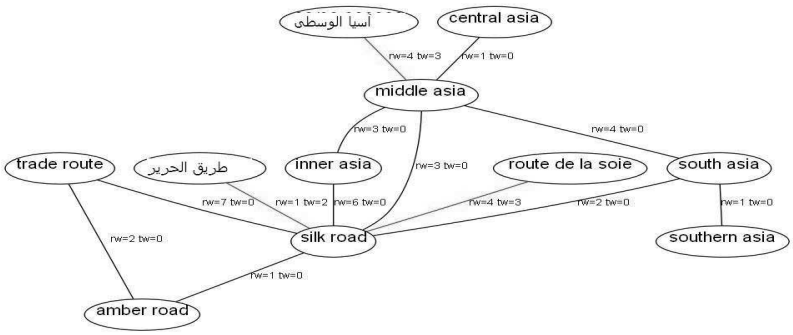

Fig. 1. A small MPG

A multilingual preterminological graph *MPG(N,E)* is a finite nonempty set *N={n1,n2, …}* of objects called Nodes together with a set *E={e1,e2, …}* of unordered pairs of distinct nodes of *MPG* called edges. *MPG* of domain *X*, contains possible multilingual terms related to that domain connected to each other with relations. A multilingual lexical unit and its translations in different languages are represented as connected nodes with labels.

In an *MPG* the set of nodes **N** consists of *p,l, s, occ,* where *p* is the string of the term, *l* is the language, *s* is the code of the first source of the term, and *occ* is the number of occurrences. Note that *l* could be undefined. For example: *N={[silk road, en, log],[Great Wall of China, en, ,wikipedia, 5], [الصين, ar, contributorx,6]}*, here we have three nodes, 2 of them are English and one in Arabic, each term came from a different source. Note that English and Arabic terms belong to the same *N* thus, the same MPG.

An *Edge e={n, v}* is a pair of nodes adjacent in an *MPG*. An edge represents a relation between two terms represented by their nodes. The nature of the relation varies. However, edges are weighted with several weights (described below) to indicate the possible nature of this relation.

The following are the weights that label the edges on an MPG: *Relation Weights rw*: For an edge *e={[p1,l1,s1], [p2,l2,s2]}, rw* indicates that there is a relation between the preterm *p1* and *p2*. The nature of the relation could not be assumed by *rw*. *Translation Weights tw*: For an edge *e={[p1,l1,s1], [p2,l2,s2]}, tw* suggests that *p1* in language *l1* is a translation of *p2* in language *l2*. *Synonym Weights sw*: For an edge *e={[p1,l1,s1], [p2,l1,s2]}, sw* suggests that *p1* and *p2* are synonyms. Weights are measures calculated based on (1) analyzing log files, (2) terminology extraction, (3) automatic lexical translation, and (4) volunteer contribution, as we will describe in the following subsections. A *tedge* is an edge where *tw* is more than zero, *tdegree(n)* is the number of *tedges* that connect to n. A *redge* is an edge where *rw* is more than zero, *rdegree(n)* is the number of *regdes* that connect to n.

Figure 1 shows a simple MPG. The shown nodes represents terms related to the historical Silk Road [10]. For example, “inner Asia” and “middle Asia” are synonyms, so *rw* between them is 3 while *tw* equals zero. “Route de la soie” is the French equivalent of “Silk Road”; hence *tw* is more than 1.

2.2 Implicit Approach

Access log files constitute a very useful resource that is related to a specific domain, as they register the interactions between a domain-related online community on one side and users (who might include domain experts) on the other side. A server access log file keeps track of the requests that have been made to the server, along with other information like request time, IP address, referred page.

We analyze two kinds of requests that can provide us with information to enrich the MPG: (1) requests made to a local search engine devoted to a website and its documents, and (2) requests with reference from a web-based search engine (like Google, Yahoo!…).

From these requests we can obtain the search terms that visitors have used to access the website. Moreover, we can understand the way users interpret a concept into lexical units. Finding a pattern in their requests may lead to find a relation between the terms used in requests. For example, if we find that five different users send two consecutive search requests t1 and t2, then there is a possibility that t1 and t2 have a lexical relation.

As the pseudo code of "*analysing_searchlogfiles()*" illustrates we construct the initial MPG from access log files after filtering their records to find the search requests. The graph constructor analyzes the requests to make the initial graph by creating edges between terms available within the same session. The relation weight between x and y, *rw(x,y)*, is set to the number of sessions containing x and y within the log file. For example, *rw(x,y)* = 10 means that 10 people thought about x and y within the same search session.

```
analysing_searchlogfiles()
    for each search session of the log file Session_n
        for each term_n in Session_i
            if there is an edge between term_i and term_j then
                rw(term_i, and term_j)++;
            elseif there is no edge
                construct edge(term_i, term_j);
                rw(term_i, term_j)=1;
```

Figure 2 shows an example of a log file and the produced graph. The proposed method did not discover the kind of relation between the terms. However it discovered that there is a relation, for example, three users requested results for "yang" followed by "yin" within the same session. Hence, edge with weight of 2 was constructed based on this fact.

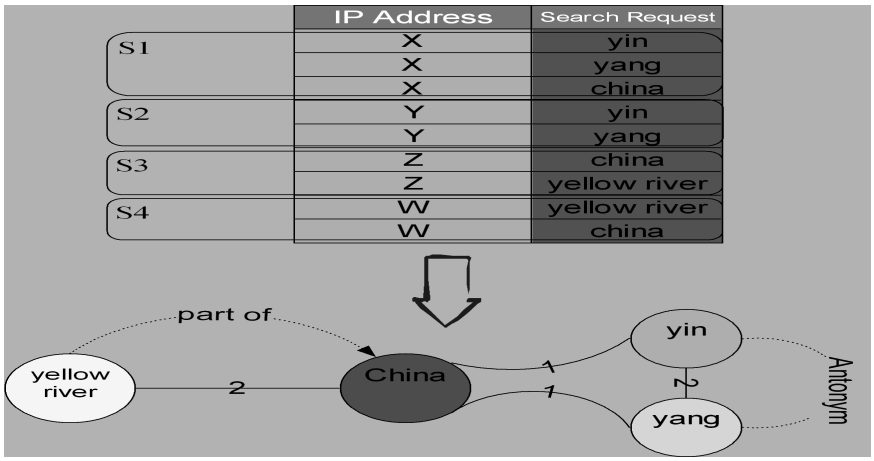

Fig. 2. Example of constructing an MPG from an access log file

2.3 Explicit Approach

Explicit user interaction is the intentional contribution from the community user. The motive to contribute is initiated by associating the contribution process to an interesting and attractive activity.

User contribution *C(x,y,s,t)* will increase the confidence that *x* is a translation of *y* s is the source language and t is the target language. Hence, *tw(x,y)* (initialized at the multililingualization process) will be increased, accordingly. Accumulating the contributions will result in enlarging the graph and enhancing the confidence in its translation equivalences.

3 Graph Multilingualization

After constructing the initial MPG, we expand it by translating the terms using multilingual online resources. In the case of terminology, we are using Wikipedia [11], IATE, and Google Translate [12]. The choice of Wikipedia comes from the fact that it could be helpful for cultural terminology [13], as it is rich with proper names.

Each translation into each language is represented as a new node in the graph, an edge between the term and its translation is established, and *tw (initially equals 0)* is modified accordingly.

The translation weight between x and y equals the number of resources indicating that x is a translation of y.

Therefore, if $tw(x,y) > k$, where k is a confidence threshold, then x is a *direct translation* equivalence of y.

More generally, if there is a path p between x and y, where all edges on p have *tw* > k, then x is an *indirect translation* (see next subsections) of y.

Based on figure 1, if k=1, then "آسیا الوسطی" is an Arabic equivalent of "middle Asia" because *tw* is larger than 1.

3.1 Finding Synonyms

There are two kinds of information to indicate that two terms may be synonym in a graph. The first one is the *rw* between both of them, and the second is the number of translations overlapping the two terms. The second information was used by TransGraph to resolve word sense inflation. However, in a domain-dedicated terminology, the first information is very important to find that some terms represent the same concept.

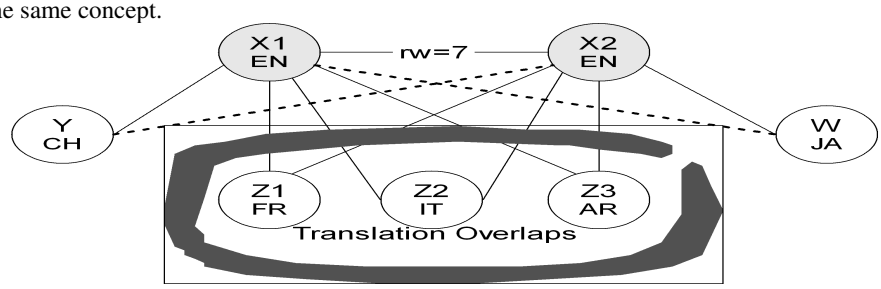

Fig. 3. An MPG where there is a possibility that X1 and X2 are synonyms

Figure 3 shows that X1 and X2 have a relation *rw* and they have three shared translations. The probability that X1 and X2 are synonym increases if the number of translation overlaps is high, based on [8], and if the *rw(X1,X2)* is high, therefore we need to find a modified weight for synonyms.

The following formula computes the new weights:

$$synonym\,weight(X1, X2) = \frac{(rw(X1, X2))}{(min(rdegree(X1), rdegree(X2)))} + \frac{(\#\,translation\,overlaps)}{(min(tdegree(X1), tdegree(X2)))} \qquad (1)$$

Where *rdegree(x)* is the number of edges that connect to x with *rw* > 0, and *tdegree(x)* is the number of edges that connect to x with *tw* > 0. For example, in figure 3, *sw(X1, X2)=(1/1+3/4)=7/4.*

3.2 Indirect Translations

If the graph has a term *t1* and its synonym *t2*, then edges with high *tw* can connect *t1* to *t2* and vice versa. In other words, *t1* and *t2* correspond to the same concept and it is

possible that they have the same translations. We call this *indirect translation* because there is no direct edge between the term and translationally equivalent terms.

Therefore, x is an indirect translation of y, if *x* connects to *x1*, and there is a path *p* from *x1* to *y*, where *tw(x1,x)* > *k*, *k* being a confidence threshold, and where all edges of *p* have *sw* > *k1*, *k1* being a synonymy confidence threshold.

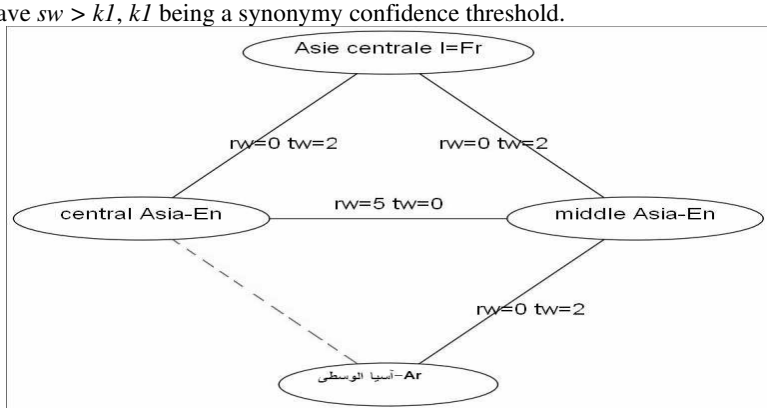

Fig. 4. An MPG with indirect translation

For example, in figure 4, "آسيا الوسطى" is considered as an Arabic translation of "central Asia" if *k=1*, and *k1=1*. This is because *tw("آسيا الوسطى", middle Asia)=2*, *and sw(central Asia, middle Asia)=5/1+1=6*, based on formula 1.

4 Platform

As figure 5 shows, the terminological lexical sphere for a domain, is constructed from different resources. And it is represented as an MPG, the MPG is used as it is for several applications that will serve the online community and the same applications are capable of attracting contribution.

The produced graph is represented as a GraphML file (graphml.graphdrawing.org). GraphML offers a structure that is compatible with MPG, and it can be easily produced by the system, adopting this format is important to make the graph more scalable and useful for other applications, beside many systems and tools have been developed to manipulate and visualize graphs in GraphML format.

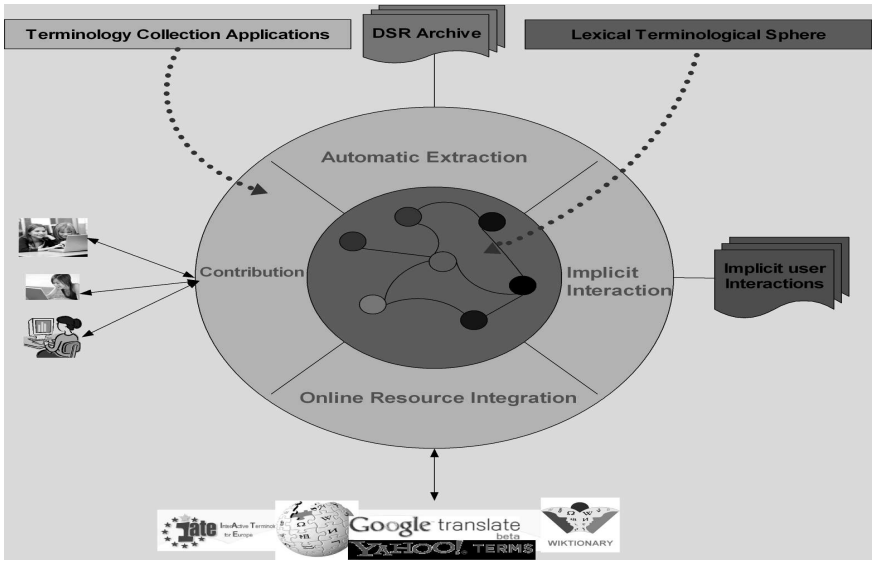

Fig. 5. Platform for constructing MPG

4.1 The Case of the Digital Silk Road

The Digital Silk Road project [14] is an initiative started by the National Institute of Informatics (Tokyo) in 2002, to archive cultural historical resources along the Silk Road, by digitizing them and making them available and accessible online.

One of the most important sub-projects is the Digital Archive of Toyo Bunko Rare Books [15] where tens of old rare books available at Toyo Bunko library have been digitized using OCR (Optical Character Recognition) technology. The digitized collection contains books from different languages (English, French, Russian…), all of them related to the historical Silk Road, like the 2 volumes of the Ancient Khotan by Marc Aurel Stein.

We are trying to build a collaborative multilingual terminological database dedicated to the DSR project and its resources [16]. To conduct such a study, there are two approaches, implicit and explicit, as described in [17]. We used the implicit approach as we have the access log files of the website since 2003, which contain many interesting facts.

DSR-MPG is synchronized with a multilingual pre- terminological database pTMDB that interacts with users who search the data base and contribute.

Each term represented as a node in the graph corresponds to a record in the Solr-based index along with some useful term related and concept related information.

For a historical archive like the DSR, we find that reading and searching were the most important for users. Log files since 2003 show that 80% of the project visitors were interested in reading the historical records. Moreover, around 140000 search

requests have been sent to the internal search engine. We are trying to derive indirect motivation to the pTMDB through the interesting resources of the DSR itself. So we implemented two applications (1) "contribute-while-reading" and (2) "contribute-while-searching", explained in the next subsection. They are available at http://dsr.nii.ac.jp/pTMDB/

4.2 Applications

4.2.1 Contribute While Searching

As shown in figure 6, historical physical books have been digitized and indexed into a SOLR-based search engine.

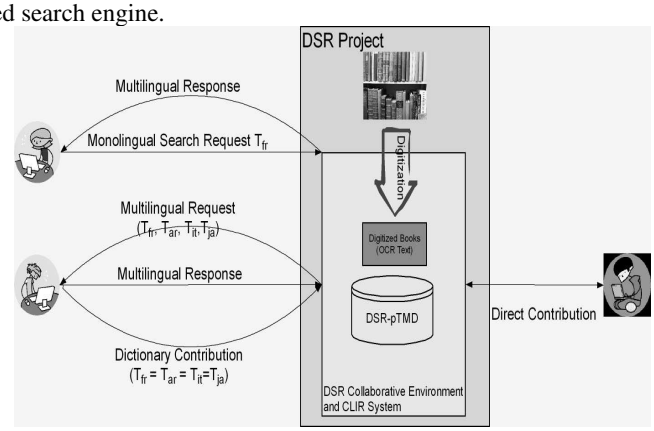

Fig. 6. General architecture of the environment [16]

We expect users to send monolingual search requests in any language supported by our system to get multilingual answers. Having a term base of multilingual equivalences could achieve this [18] [19]. A bilingual user who could send a bilingual search request could be a valid candidate to contribute. In fact, the same bilingual request could be a valid MPG contribution, and also multilingual requests. We plan that users who use our search engine will use the DSR-pTMDB to translate their requests and will contribute to the graph spontaneously.

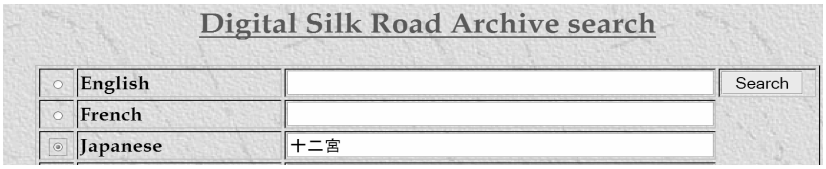

Fig. 7. A Japanese user translating his request

As figure 7 shows, a Japanese user would translate the search request, to receive the results, as shown in figure 8.

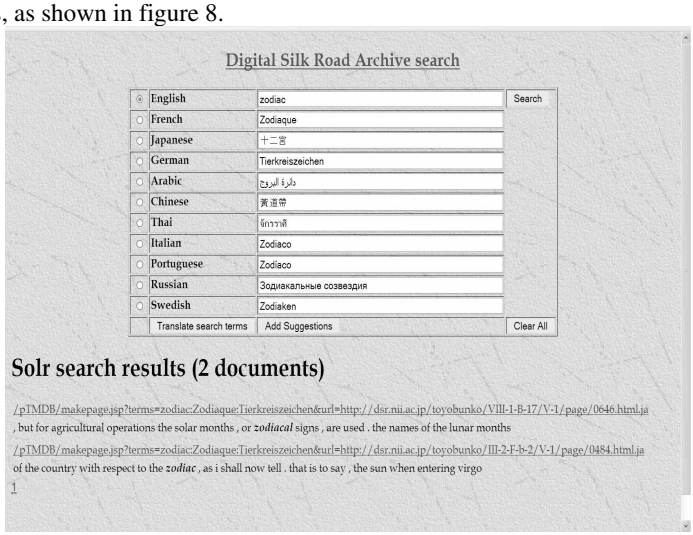

Fig. 8. Search results

During the searching process, the user can ask to add new translation if s/he was not happy with the suggested translation, by clicking on “Add Suggestions” to view the page showed at figure 9.

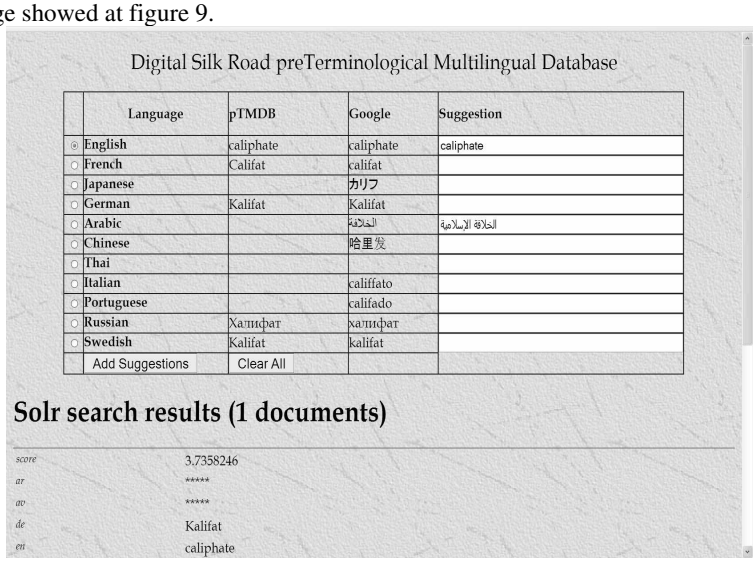

Fig. 9. Contribution page

4.2.2 Contribute While Reading

The other interesting application is trying to help users from different linguistic backgrounds to translate some of the difficult terms into their languages while they are reading, simply by selecting a term from the screen.

As shown in figure 10, readers will see a page from a book as an image, with its OCR text. Important terms will be presented with yellow background. Once a term is clicked, a small child contribution/lookup window will be open, similar to the one in figure 9. Also user can lookup/translate any term from the screen by selecting it.

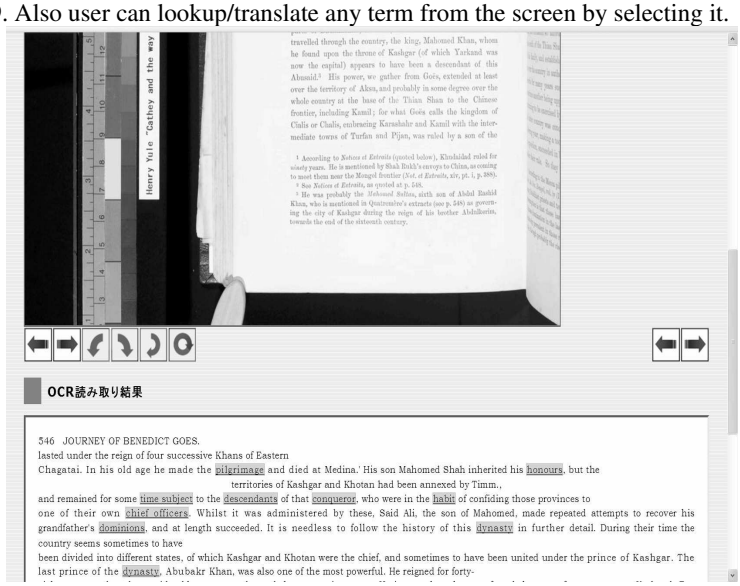

Fig. 10. Translate while reading

This application helps covering all the important terms of each book.

5 Experiment

To build the initial DSR-MPG, we used the access log files of the DSR website (dsr.nii.ac.jp) from December 2003 to January 2009. The initial graph after normalization contained 89,076 nodes; most of them being for English terms, we filtered the logs (semi automatically) to analyze only access requests with search queries, the initial graph was produced in less than 5 hours, using a PC with Intel Pentium 4 processor. We also sent the OCR text of the archived books of Toyo Bunko library to a term extraction engine, in this experiment Yahoo! Terms. We extracted 81,204 terms. 27,500 of them were new terms that were not discovered from the

access log files. So, the total number of nodes in the initial graph was 116,576 nodes, see figure 11 for sample nodes.

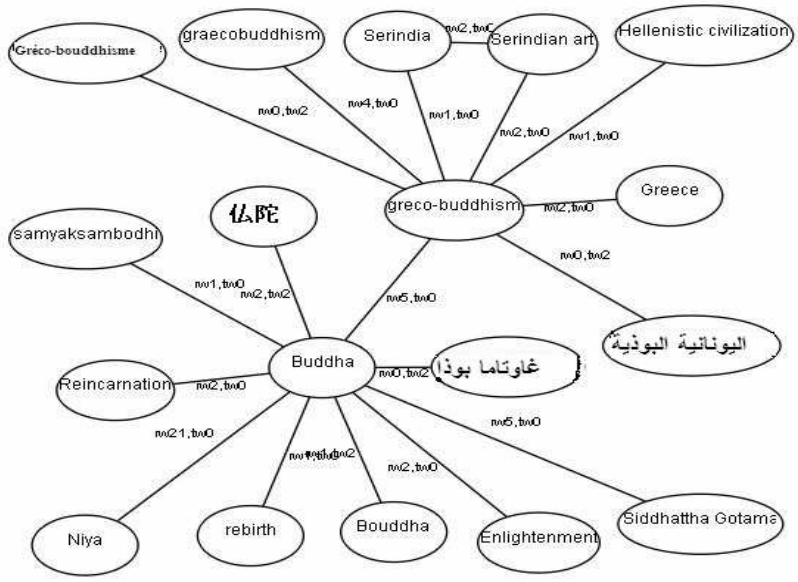

Fig. 11. Sample nodes from DSR-MPG

After the multilingualization process described in section 3, the graph has 210,781 nodes containing terms from the most important languages. The graph has now 779,765 edges with $tw > 0$.

The important languages are the languages of the majority of the visitors, the languages of the archived books, and representative languages a long the Silk Road. DSR-MPG achieved high linguistic coverage as 20 languages have more than 1000 nodes on the graph.

To evaluate the produced graph, we extracted 350 English terms manually from the index pages of the following books:

- Ancient Khotan, vol.1: http://dsr.nii.ac.jp/toyobunko/VIII-5-B2-7/V-1/
- On Ancient Central-Asian Tracks, vol.1: http://dsr.nii.ac.jp/toyobunko/VIII-5-B2-19/V-1
- Memoir on Maps of Chinese Turkistan and Kansu, vol.1: http://dsr.nii.ac.jp/toyobunko/VIII-5-B2-11/V-1

We assume that the terms available in these books are strongly related to the DSR. Hence, we tried to translate them into Arabic and French.

Figure 12 compares between DSR-MPG, and various general purpose dictionaries. Out of the 350 terms, we found 189 correct direct translations into Arabic. However, the number reached 214 using indirect translations.

On the other hand, the closest to our result was PanImages, which uses Wikitionaries [20] and various dictionaries, with only 83 correct translations.

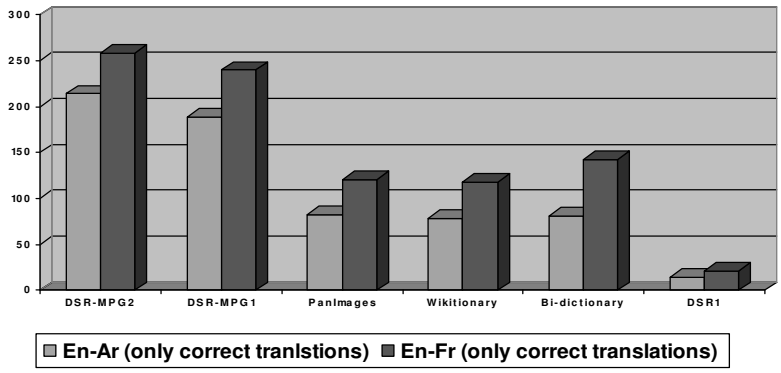

Fig. 12. A comparison between DSR-MPG, and other dictionaries. The En-Ar bi-dictionary is Babylon [21], and the En-Fr bi-dictionary was IATE

DSR-MPG1 is the translations obtained from formula 1, DSR-MPG2 represents the translations obtained from indirect translations, which increased the amount of correct translation by 25 terms in the case of En-Ar.

The result can be progressively enhanced by accepting contributions from volunteers through the applications we described in the section three and the generic nature of MPG makes it easy to accept contributions from any dictionary or terminological database.

Conclusions and Perspectives

We described the explicit and implicit approaches that we are using to extract and discover domain dedicated terminology from human interaction with a related web-community. We presented a new lexical resource that can handle multilingual terms for a specialized domain. Multilingual Preterminological Graphs are constructed based on domain dedicated resources, and based on volunteer contributions. We described the approach for using access log files to initialize such graphs by finding the trends in the search requests used to access the resources of an online community.

Aiming at a standardized multilingual repository is very expensive and difficult. Instead of that, MPGs tries to use all available contributions. This way will enhance the linguistic and informational coverage, and tuning the weights (*tw*, *rw*, and *sw*) will give indications for the confidence of the translation equivalences, as the *tedges* accumulate the agreements of the contributors and MDRs (online resources).

We experimented the concept of MPGs on the domain of the historical Silk Road. We used the resources of the Digital Silk Road Project to construct a DSR-MPG and some applications that attract further contribution to the MPG. DSR-MPG achieved high linguistic and informational coverage compared to other general purpose dictionaries. Furthermore, the generic structure of the MPG makes it possible to

accept volunteer contributions, and it facilitates further study of computing more lexical functions and ontological relations between the terms.

We are packaging the platform for constructing the DSR-MPG to be used in other domains as a tool for "Multilingual Terminology Elicitation" this platform will construct an MPG for a set of textual resources to collect its preterminology. Currently we are working on constructing an MPG for the domain of Arabic dreams interpretation, this MPG will serve a service for interpreting dreams (in Arabic), a beta version is available at this website [22]. Furthermore, we are investigating more scenarios of contribution to enrich the graph; one of them is based on playful methods using a *game with a purpose (GWAP)* [23] where users contribute to a knowledge repository while playing an online game, the knowledge repository in our case is a lexical graph.

References

1. Cabré, M.T. and J.C. Sager, *Terminology: Theory, methods, and applications*. 1999: J. Benjamins Pub. Co. xii, 247 p.
2. Kageura, K., *The Dynamics of Terminology: A descriptive theory of term formation and terminological growth*. Terminology and Lexicography Research and Practice 5. 2002. 322 p.
3. IATE. *Inter-Active Terminology for Europe*. 2008 [cited 2008 10/10/2008]; Available from: http://iate.europa.eu.
4. UN. *United Nations Multilingual Terminology Database*. 2008 [cited 2008 10/10/2008]; Available from: http://unterm.un.org/.
5. IEC. *Electropedia*. 2008 [cited 2008 10/10/2008]; Available from: http://dom2.iec.ch/iev/iev.nsf/welcome?openform.
6. Gopestake, A., et al., *Acquisition of lexical translation relations from MRDS.* Machine Translation, 1994. **Volume 9, Numbers 3-4 / September, 1994**: p. 183-219.
7. Helmreich, S., L. Guthrie, and Y.A. Wilks. *The use of machine readable dictionaries in the PANGLOSS project*. in *In Proceedings of the AAAI Spring Symposium on Building Lexicons for Machine Translation (Stanford Univ.)*. 1993.
8. Etzioni, O., et al. *Lexical translation with application to image searching on the web*. in *MT Summit XI*. 2007. Copenhagen, Denmark.
9. Anh, L.V., *Human Computation*, in *Computer Science*. 2005, Carnegie Mellon University Pittsburgh. p. 87 pages.
10. Ono, K., et al. *Memory of the Silk Road -The Digital Silk Road Project-*. in *Proceedings of (VSMM08), Project Papers*. 2008. Limassol, Cyprus.
11. Wikipedia. *Wikipedia*. 2008 [cited 2008 1 June 2008]; Available from: http://www.wikipedia.org/.
12. Google. *Google Translate*. 2008 [cited 2008 1 June 2008]; Available from: http://translate.google.com.
13. Jones, G.J.F., et al. *Domain-Specific Query Translation for Multilingual Information Access Using Machine Translation Augmented With Dictionaries Mined From Wikipedia*. in *Proceedings (CLIA-2008)*. 2008. Hydrabad, India.

14. NII. *Digital Silk Road*. 2003 [cited 2008 1/9/2008]; Available from: http://dsr.nii.ac.jp/index.html.en.
15. NII. *Digital Archive of Toyo Bunko Rare Books*. 2008 [cited 2008 1 June 2008]; Available from: http://dsr.nii.ac.jp/toyobunko/.
16. Daoud, M., et al. *A CLIR-Based Collaborative Construction of Multilingual Terminological Dictionary for Cultural Resources*. in *Translating and the Computer 30*. 2008. London-UK.
17. Stermsek, G., M. Strembeck, and G. Neumann. *A User Profile Derivation Approach based on Log-File Analysis*. in *in Hamid R. Arabnia & Ray R. Hashemi, ed., 'IKE' , CSREA Press*. 2007.
18. Chen, A., *Cross-Language Retrieval Experiments at CLEF 2002*. in CLEF-2002 working notes,, 2002.
19. Oard, D., *Global Access to Multilingual Information*, in *Fourth International Workshop on Information Retrieval with Asian Languages*. 1999: Taipei-Taiwan.
20. Wikitionary. *Wikitionary*. 2008 [cited 2008 1/9/2008]; Available from: http://en.wikipedia.org/wiki/Wiktionary.
21. Babylon. *Babylon Dictionary*. 2009 [cited 2009 5/5/2009]; Available from: http://www.babylon.com/define/98/English-Arabic-Dictionary.html.
22. Daoud, D. Tafseer Al Ahlam. 2010 [cited 2010; Available from: http://www.maherinfo.com/.
23. Ahn, L.v., Games With A Purpose. IEEE Computer Magazine, 2006: p. pp 96-98.

Appendix: MPG as a GraphML

The following is a sample MPG in GraphML format:

————XML STARTS————————————

```
<?xml version=”1.0″ encoding=”UTF-8″?>
<GraphML>
<key id=”d0″ for=”node” attr.name=”preterm” attr.type=”string”></key>
<key id=”d1″ for=”node” attr.name=”language_code” attr.type=”string”>
<default>eng</default></key>
<key id=”d2″ for=”node” attr.name=”source” attr.type=”string”>
<default>unknown</default></key>
<key id=”d3″ for=”node” attr.name=”occ” attr.type=”string”>
<default>1</default></key>
<key id=”d4″ for=”edge” attr.name=”rw” attr.type=”double”>
<default>0</default></key>
<key id=”d5″ for=”edge” attr.name=”sw” attr.type=”double”>
<default>0</default></key>
<key id=”d6″ for=”edge” attr.name=”tw” attr.type=”double”>
<default>0</default></key>

<graph isAcyclic=”true” id=”dsr” >
<node id=”2353″ >
<data key=”d0″>great wall of China</data>
<data key=”d1″>eng</data>
<data key=”d2″>dsr_log</data>
<data key=”d3″>11</data>
```

</node>
<node id=”2354″ >
<data key=”d0″>سور الصين العظيم</data>
<data key=”d1″>ara</data>
<data key=”d2″>wikipedia</data>
<data key=”d3″>3</data>
</node>
<node id=”2355″ >
<data key=”d0″>万里の長城</data>
<data key=”d1″>jpn</data>
<data key=”d2″>dsr_log</data>
<data key=”d3″>4</data>
</node>
<node id=”2356″ >
<data key=”d0″>Grande Muraille de Chine</data>
<data key=”d1″>fra</data>
<data key=”d2″>wikipedia</data>
<data key=”d3″>3</data>
</node>

<edge source=”2353″ target=”2354″>
<data key=”d4″>0</data>
<data key=”d5″>0</data>
<data key=”d6″>3</data>
</edge>
<edge source=”2353″ target=”2355″>
<data key=”d4″>1</data>
<data key=”d5″>0</data>
<data key=”d6″>2</data>
</edge>
<edge source=”2353″ target=”2356″>
<data key=”d4″>0</data>
<data key=”d5″>0</data>
<data key=”d6″>2</data>
</edge>
</graph>
</GraphML>
————————-XML ENDS————————-

The above XML graph corresponds to the graph in Figure 13.

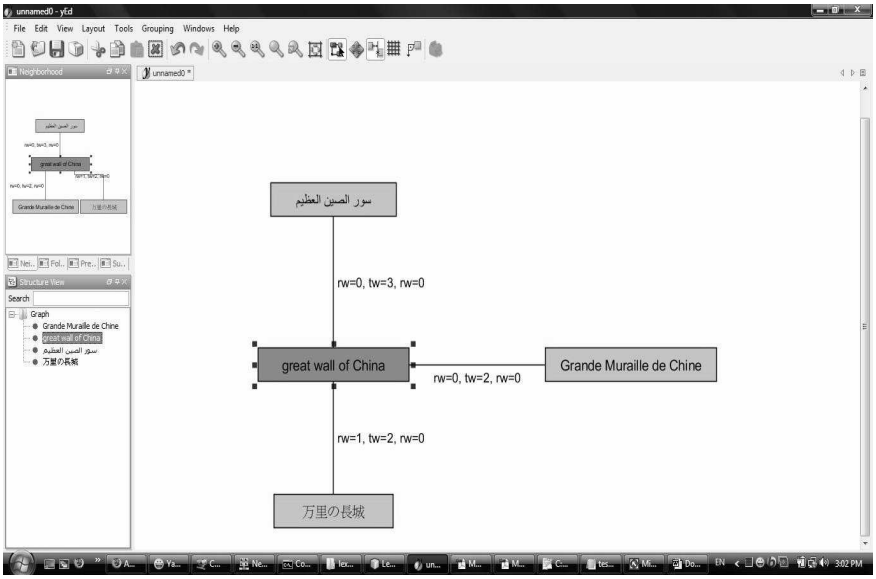

Fig. 13. Sample graph