Multi-user Feedback for Large-scale Cross-lingual Ontology Matching

Mamoun Abu Helou[1] and Matteo Palmonari[2]
[1]*Department of Management Information Systems, Al-Istiqlal University, Palestine*
[2]*Department of Informatics, Systems and Communication, University of Milan-Bicocca, Italy*

Keywords: Users Feedback, Interactive Mapping, Cross-Lingual Ontology Mapping.

Abstract: Automatic matching systems are introduced to reduce the manual workload of users that need to align two ontologies by finding potential mappings and determining which ones should be included in a final alignment. Mappings found by fully automatic matching systems are neither correct nor complete when compared to gold standards. In addition, automatic matching systems may not be able to decide which one, among a set of candidate target concepts, is the best match for a source concept based on the available evidence. To handle the above mentioned problems, we present an interactive mapping Web tool named ICLM (Interactive Cross-lingual Mapping), which aims to improve an alignment computed by an automatic matching system by incorporating the feedback of multiple users. Users are asked to validate mappings computed by the automatic matching system by selecting the best match among a set of candidates, i.e., by performing a mapping selection task. ICLM tries to reduce users' effort required to validate mappings. ICLM distributes the mapping selection tasks to users based on the tasks' difficulty, which is estimated by considering the lexical characterization of the ontology concepts, and the confidence of automatic matching algorithms. Accordingly, ICLM estimates the effort (number of users) needed to validate the mappings. An experiment with several users involved in the alignment of large lexical ontologies is discussed in the paper, where different strategies for distributing the workload among the users are evaluated. Experimental results show that ICLM significantly improves the accuracy of the final alignment using the strategies proposed to balance and reduce the user workload.

1 INTRODUCTION

With the emergence of the Semantic Web vision, the Web has witnessed an enormous growth in the amount of multilingual data that exist in a large number of resources. Since then, there has been an increasing interest in accessing and integrating these *multilingual resources* (Hovy et al., 2012). *Ontologies* have been proposed for the ease of data exchange and integration across applications. When data sources using different ontologies have to be integrated, mappings between the concepts described in these ontologies have to be established. This task is also called *ontology mapping*. Ontology mapping methods perform two main sub tasks: in *candidate match retrieval*, a first set of potential matches is found; in *mapping selection*, a subset of the potential matches is included in a final alignment.

The problem of finding mappings between concepts lexicalized in different languages has been addressed in the field of *Cross-lingual Ontology Mapping* (Spohr et al., 2011). Cross-lingual ontology mapping is currently considered an important challenge (Garcia et al., 2012), which plays a fundamental role in establishing semantic relations between concepts lexicalized in different languages, in order to align two language-based resources (Trojahn et al., 2014), create multilingual lexical resources with rich lexicalizations (Navigli and Ponzetto, 2012), or support bilingual data annotation (Zhang, 2014). Automatic matching systems are introduced to ease this task by finding potential mappings and determining which ones should be included in a final alignment. Automatic cross-lingual mapping methods can be used either to compute mappings automatically (de Melo and Weikum, 2012), even at the price of accuracy, or to support semi-automatic mapping workflows by recommending mappings to users (Pazienza and Stellato, 2006).

In a recent work, we define a lexical similarity measure based on evidence collected from translation resources and we run a local similarity optimization algorithm to improve the assignments between source and target concepts (Abu Helou and Palmonari, 2015). In particular, we define **selection task** as the task of selecting the correct target con-

Abu Helou M. and Palmonari M.
Multi-user Feedback for Large-scale Cross-lingual Ontology Matching.
DOI: 10.5220/0006503200570066
In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD 2017)*, pages 57-66
ISBN: 978-989-758-272-1

cept, which one source concept should be mapped to among a set of candidates ranked by similarity. The selection of a correct mapping from a set of candidate matches still remains a difficult task, in particular when *contextual knowledge* is limited or cannot be used to disambiguate the meaning of the concepts. For instance, mappings involving lexicalized concepts by only one word (synonymless), which has several meaning (polysemous), e.g., the concept {table}(defined in Table 1), are harder to filter out within the mapping selection task (Abu Helou et al., 2016). With automatic matching systems, different candidates may be evaluated as equally good matches for a source concept based on the available evidence, i.e., a *tie* occurs among a set of top-ranked matches. In this case, the mapping for this source concept is *undecidable*, and no mapping for this concept is included in the final alignment (otherwise, we say that the mapping is *decidable*). Resolving ties by randomly selecting one of the highest ranked candidate matches comes at the price of precision. Otherwise, no mapping for this concept can be included in the final alignment at the price of recall.

This paper investigates, in the cross-lingual ontology matching domain, the adoption of semi-automatic matching methods where multiple users are involved in the mapping selection processes. Web applications have been proposed to help individual users with difficult cross-lingual matching tasks, as the task of linking short service descriptions lexicalized in different languages (Narducci et al., 2017). Beyond this, interactive matching processes that aggregate feedback provided by multiple users, either experts (Cruz et al., 2014) or crowd workers (Sarasua et al., 2012) seem particularly promising for large scale cross-lingual matching tasks. For instance, if a correct match can be found among a set of top-ranked candidate matches, and if this set is reasonably small, one could use interactive mapping approaches to let users decide about the mappings. Early experiments conducted in previous work, in which such scenarios were investigated, showed potential improvement in recall (Abu Helou and Palmonari, 2015).

This paper presents ICLM (Interactive Cross-lingual Mapping) application, which is a semi-automatic matching approach that supports feedback provided by multiple users. In the approach proposed in this paper, an alignment is first computed using automatic matching methods. Then, users are asked to establish mappings for a subset of source concepts that are estimated to require the user feedback to be mapped. In particular, the same selection task is assigned to more users to collect consensus over the decision. Users can also decide the type of relation defined by the mapping between the source and the target concepts (equivalent to, more specific than, or more general than, as explained in detail in Section 5). We define a new strategy to select the mappings that are worth being validated by the users, based on the evaluation of the difficulty of the selection tasks. This evaluation is based on the lexical characterization of concepts under evaluation, i.e., on the estimation of the ambiguity conveyed by the concepts involved in mappings (Abu Helou et al., 2016), under the hypothesis that most difficult selection tasks require the agreement of more users. Using the same principle, we estimate the number of users that are asked to perform a selection task, which determines the overall user effort consumed to decide upon a selection task. Experimental results show that the proposed interactive matching method, with dynamic selection of selection tasks on which the user feedback is required and the dynamic allocation of user effort, improves significantly the quality of the final alignment both in terms of precision and recall.

The rest of this paper is organized as follows. Section 2 overviews related work. Section 3 overviews ICLM and provides more insights on the functionalities provided by the approach and on the interface provided to users. Further, Section 4 describes the key elements of the proposed approach: strategies to estimate the validation efforts required from users for each source concept. In Section 5, we discuss the conducted experiment: the dataset, a model for evaluating the quality of the mappings and users effort, and the results. Finally, in Section 6, we draw some conclusions and describe future work.

2 RELATED WORK

In this section we review related work on engaging users feedback in matching processes, and highlight the role of concept lexicalizations in estimating the selection tasks' difficulties in the cross-lingual matching domain.

Since the performance of automatic matching systems is limited; leveraging the contribution of the user feedback has been recognized as a fundamental step to validate candidate matches. Semi-automatic mapping workflows has been adopted in several data integration systems, including ontology matching systems; either by collecting feedback given by a single user or multiple users, to support validation processes.

Approaches designed for a single-user scenario are developed first. Some heuristics are used to support the user in building the MultiWordNet, a multilingual lexical ontologies (Pianta et al., 2002), by suggesting a set of potential mappings. An entity linking web application CroSeR developed to support the cross-language linking of e-Government services to the Linked Open Data cloud (Narducci et al., 2013; Narducci et al., 2017). The user can select a service in a source catalog and use the ranked list of matches suggested by CroSeR to select the equivalent service. The approach may be used to find more candidate target concepts; however, in this paper we focus more on the problem of mapping selection; so this is left for future work. Moreover, several research works used the single-user scenario; including works proposed in (Noy and Musen, 2003; Shi et al., 2009; Cruz et al., 2012; Jimnez-Ruiz et al., 2012). However, in this paper we focus on the problem of involving multiple users in the mapping selection tasks.

On the other hand, multi-user feedback scenarios are also used (Cruz et al., 2014; Sarasua et al., 2012; Demartini et al., 2012). Crowdsourcing approach is used to collect feedback from multiple users (called workers) for individual candidate mappings. They include CrowdMap (Sarasua et al., 2012) for ontology matching, ZenCrowd (Demartini et al., 2012) for entity linking. Crowdsourcing based systems assign the same number of users for every task; in this paper we investigate a controlled strategy for dynamically determine the number of users required for each matching task. We distribute the mapping selection tasks over some users based on the difficulty of the selection tasks. In our approach, similar to CrowdMap, *consensus* is obtained on the mappings before they are included in the final alignment. Mappings in ZenCrowd are considered correct if they have a posterior probability that is greater than a threshold. A pay-as-you-go approach in which the alignment is refined after each iteration is used in (Cruz et al., 2014; Cruz et al., 2016), in which they adopted a feedback propagation strategy; at each iteration the alignment is recomputed using the full mapping space, which makes it unfeasible when considering large ontologies as the ones considered in this paper. In addition, since we are focusing on mappings between lexical ontologies, we use insights about lexical ambiguity gained in previous work (Abu Helou and Palmonari, 2015; Abu Helou et al., 2016), in which the *Local Similarity Optimization Algorithm* (LSOA) is introduced. *LSOA* automatically selects the mappings based on merging locally optimal assignments computed for each source concept. In this paper we consider *LSOA* focusing on cross-lingual mapping scenarios where lexically-rich resources are not structured and to leverage the concepts' lexicalization to estimate the selection tasks difficulties. However, structural matching methods (Shvaiko and Euzenat, 2013; Trojahn et al., 2014; Cruz et al., 2009) can be easily incorporated in the similarity evaluation step without major changes to our approach.

We find that, the observations derived from studying the difficulty of the mapping selection tasks (Abu Helou et al., 2016) are particularly useful for similar approaches, because they can help to decide on which mappings the user inputs are more valuable. In particular, when we consider the undecidable mappings. A large-scale study, which include cross-lingual mappings from large lexical ontologies in four different languages, is conducted on the effectiveness of automatic translation resources on cross-lingual matching (Abu Helou et al., 2016). Concepts (or, synsets: sets of synonym words) are classified based on different lexical characteristics: word ambiguity (e.g., monosemous vs polysemous), number of synonyms (e.g., synonymful vs synonymless), and position in a concept hierarchy (e.g., leaves vs intermediate concepts). Table 1 summarizes the lexical classification of concepts. Using these classifications, the effectiveness of automatic translations is evaluated by studying the performance on the cross-lingual mapping tasks executed using automatic translations for different categories of concepts. Evidence collected from automatic translations is used in a baseline mapping selection approach, i.e., majority voting, to evaluate the difficulty of the mapping selection task. The study reveals several observations; for instance, for synonymful concepts, the larger the number of synonym words covered by translations, the easier the mapping selection task is. While mapping involving polysemous but synonymless concepts (*P&OWS*) are harder to filter out within mapping selection task; thus we need to collect more evidence, or to involve (more) users, to select the correct mappings.

For rich studies on involving users during the matching processes, and on estimating the selection task difficulties; we refer to the work of (Dragisic et al., 2016) and (Abu Helou et al., 2016), respectively.

3 ICLM OVERVIEW

ICLM[1], Interactive Cross-Lingual Mapping, is a Web application that supports a semi-automatic mapping procedure aiming at speeding up and improving an automatically generated alignment. ICLM tries to reduce the users' efforts in validating cross-lingual concept mappings. Figure 1 shows the home page of ICLM. ICLM distributes the mapping selection tasks

[1]http://193.204.59.21:1982/iclm/

Table 1: Concepts (synsets) categories.

Category	Synset name	Definition "*synsets that have...*"
OWS	*One-Word*	only one word (*synonymless synset*).
MWS	*Many-Words*	two or more synonym words (*synonymful synset*).
M&OWS	*Monosemous and OWS*	only one word, which is also a monosemous word (e.g., {desk}).
M&MWS	*Monosemous and MWS*	two or more synonym words, which are all monosemous words (e.g, {tourism, touristy}).
MIX	*MIXed*	monosemous and polysemous synonym words (e.g., {table, tabular array}).
P&OWS	*Polysemous and OWS*	only one word, which is also a polysemous word (e.g., {table}).
P&MWS	*Polysemous and MWS*	two or more synonym words, which are polysemous words (e.g,{board, table}).

Figure 1: ICLM home page.

on users based on the mapping tasks' difficulties, i.e., in the selection process, ICLM defines the number of users based on the difficulty of the mapping selection task. ICLM estimates the difficulties of the mapping selection tasks based on lexical characteristics (explained in Section 4) of concepts under evaluation and on how confident the automatic matching algorithm is, i.e., ICLM estimates the selection task difficulty and accordingly estimates the expected users effort (number of users to validate a mapping task).

Initially the source concepts will be automatically matched against the target concepts using automatic matching methods. Then, the system estimates the mapping selection tasks difficulty; and accordingly defines the number of users to validate each task (explained in details in Section 4). In this way, ICLM distributes the mapping tasks over some users based on the estimated efforts of selection tasks, unlike pure crowdsourcing models, e.g., (Sarasua et al., 2012), which equally assign the same number of users for every task. The user is free to select any source concept from the source list. Once the user identifies the potentially correct candidate match, he can choose one relationship that reflects his decision (described in details in Section 5).

Since more than one user is involved, ICLM uses a *consensus-based* approach to decide whether a mapping belongs to the final alignment. Similar to previous work (Cruz et al., 2014), ICLM uses a consensus model based on simple majority vote, where V is an odd number of validations considered sufficient to decide by majority (ICLM does not require that all the users validate each mapping task); thus, *minimum consensus*, $\mu = \lfloor (V/2) + 1 \rfloor$, is the minimum number of similar vote that is needed to make a final decision on a mapping. For example, if $V = 5$ is the number of validations considered sufficient to decide by majority, a final decision on a mapping can be taken when $\mu = 3$ similar vote are assigned to a mapping by the users. Every mapping that obtains the minimum consensus of votes will be confirmed, i.e., included in the final alignment, and will be removed from the source concepts list for this specific user. Once the user finishes his task, a confirmation message is sent, and the corresponding task is removed from the source list. However, other users may still find it, for instance, if the minimum consensus of votes has not reached. After each validation task, ICLM updates the source list until the whole mappings are validated. In this way, ICLM reduces and saves more of users efforts. Cases where agreement (the minimum consensus of votes) is not achieved, the match which has the highest rank and received more votes will be included in the final alignment. Otherwise, it will not be included in the final alignment.

Observe that the agreement factor (i.e., the minimum consensus of votes) can be tuned in a favor to increase the mapping accuracy by increasing this factor. However, this comes at a price of increasing the users effort. Furthermore, different agreement strategies can be adopted. For example, mapping tasks will be confirmed only if a given number of users have agreed without controlling the number of users who are validating the mapping tasks. This of course will increase the users efforts. One may consider feedback reconciliation models more sophisticated than majority or weighted majority voting, for example, tournament solutions (Cruz et al., 2014). This would be an interesting direction as a future work to explore.

Next, we provide more insight on the functionali-

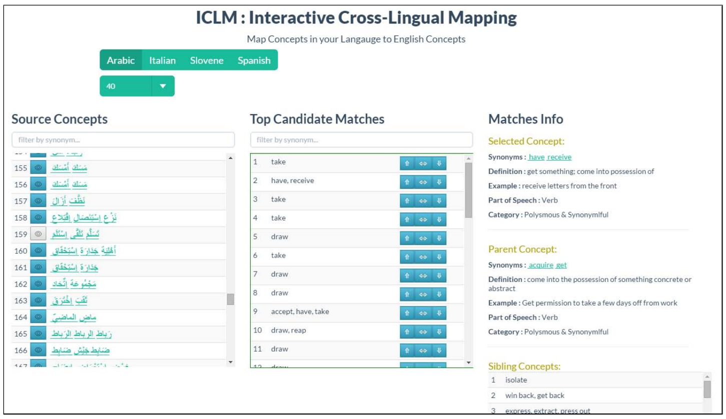

Figure 2: ICLM: supports user with useful details.

ties provided by the application and on the Web GUI[2]. Figure 2 illustrates ICLM's functionalities. Before the users start using the application, the source concepts (e.g., in Arabic) are automatically matched to the English target concepts using a lexical based disambiguation algorithm; the Translation-based Similarity Measure with Local Similarity Optimization Algorithm ($TSM + LSOA$) (Abu Helou and Palmonari, 2015). The first step that the ICLM user should perform is to *Register* and *Login*, so to enable all the validation functionalities. After that, he select the respective language of concepts to be mapped to concepts in the English WordNet (Fellbaum, 1998). The user is now able to explore the source concepts by scrolling the whole list of concepts (*Source Concepts*) or by performing a keyword-based search (see Figure 2). Next, the user selects a source concepts and ICLM retrieves a list of candidate English concepts (*Top Candidate Matches*), that are potentially equivalent. The number of retrieved matches is configurable by the user through the GUI (e.g., the 40 top-ranked matches).

Since, the connection between the source concept and the target concept could be not straightforward by simply comparing concepts' lexicalization, a user can then select a candidate match and look at further details (*Matches Info*) directly gathered from the English WordNet. Moreover, the user can click the source and target concepts lexicalization to get further information, as depicted in Figure 3. For instance, the user will be able to access an online glossary for the source language[3], as well as navigate through the semantic hierarchy of the English WordNet via the online English WordNet website[4]. Finally, the user can switch on the feedback mode of ICLM which would store the selected relation between the source concept and the English concept. For each mapping selection task ICLM logs the users' activities: the elapsed time of each mapping selection task; and users' navigation activities (accessing the external resources: the glossary or the online English WordNet). In this way, we can evaluate the effectiveness and usability of ICLM, discussed in Section 5.

4 ESTIMATING THE DIFFICULTY OF THE SELECTION TASKS

The basic idea behind ICLM is to reduce the users' effort in validating a pre-defined alignment, and thus speeding up the mapping process. In order to estimate the mapping selection tasks' difficulties, so as to estimate the required efforts (number of users required for validation), ICLM leverages the lexical characteristics of concepts under evaluation, where the confidence of the candidate matches is based on the lexical based disambiguation algorithm $TSM + LSOA$ (Abu Helou and Palmonari, 2015).

ICLM considers the following features of concepts under evaluation to estimate the validation tasks difficulties, Table 1 illustrates some examples:

- Ambiguity of lexicalization:
  - Monosemous words: words that have only one sense (meaning), e.g., the word "tourism" is a monosemous.
  - Polysemous words: words that have two or more senses, e.g., the word "table" is a polysemous.

[2]ICLM Web GUI has been adapted from CroSeR Web GUI (Narducci et al., 2013).

[3]In the current implementation, for Arabic ICLM uses Al-maany glossary: http://www.almaany.com/ar/dict/ar-ar/, which returns all possible senses of a given word (i.e, the word is not disambiguated).

[4]http://wordnetweb.princeton.edu/perl/webwn

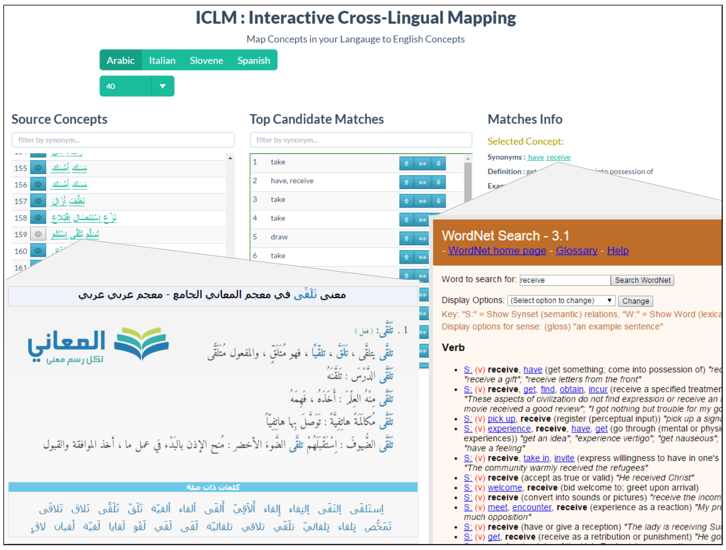

Figure 3: ICLM's details snapshot.

Polysemous words are more difficult to disambiguate than the monosemous words when contextual knowledge is limited.

- Synonym-richness:
  - Synonymless: a concept that is lexicalized with one word, e.g., the concept {table}.
  - Synonymful: a concept that is lexicalized with many words, e.g., the concept {board, table}. The more the coverage for synonym words (in synonymful synsets), the easier is the mapping selection task.
- Uncertainty in the selection Step: matches which can be obtained by an automatic cross-lingual mapping systems, in which the candidate matches are ranked based on their similarity degree with the source concepts.
  - TopOne: if there exists a unique top-ranked candidate match for the source synset.
  - TopSet: if there exists a tie (a set of top-ranked matches), i.e., a unique top-voted candidate match does not exist for the source synset.

Three **validation strategies** have been defined in ICLM: Low, Mid, and High levels of difficulty. Respectively, in each level ($l := \{L, M, H\}$), different number of users are required to validate the mapping of each source concept, i.e., the number of mapping selection tasks that are considered sufficient to decide by majority ($V^l \geq 0$). The validation strategies levels are as follow:

- **Low-difficulty:** V^L validation tasks are required.
- **Mid-difficulty:** V^M validation tasks are required.
- **High-difficulty:** V^H validation tasks are required.

Each level can have a different agreement factor, i.e., the minimum consensus of votes. Accordingly, different configurations can be considered as trade-offs between mappings accuracy and users efforts. For instances, $V^l = 0$ suggests that mappings will be directly included in the final alignment without any feedback (validation). An increase in the value of V^l means increasing the users efforts and the mapping accuracy, under the assumption that users are expected to identify the correct relation with out introducing errors (which is not always the case (Cruz et al., 2014)). Observe that, more validation strategy levels can be introduced, based on application requirements.

ICLM applies the following rules in order to select the respective validation strategy, i.e., define the number of validation tasks that are considered sufficient to decide by majority:

- **Low-difficulty:** if a monosemous synset is under evaluation and TopOne candidate match exist, *OR* if a synonymfull synset is under evaluation and TopOne candidate match exist.
- **Mid-difficulty:** if a source synset does not have a TopOne match.
- **High-difficulty:** if a source synset is polysemous and synonymless (P&OWS).

5 EXPERIMENT

The goal of this experiment is to investigate the effectiveness of ICLM in suggesting good candidate matches; not only for equivalent relation but also

Table 2: Sample dataset: distribution by category.

	Concepts category					**Total**
	M&OWS	M&MWS	MIX	P&OWS	P&MWS	
EnWN (%)	33596 (28.6)	23819 (20.2)	18676 (15.9)	30279 (25.7)	11289 (9.6)	147306 (100.0)
ArWN (%)	1995 (19.3)	1386 (13.4)	2559 (24.7)	2194 (21.2)	2215 (21.4)	13866 (100.0)
Sample (#concepts)	48	36	62	52	52	250
#Decidable mappings	24	18	31	26	26	125
#Undecidable mappings	24	18	31	26	26	125

for relationships different from the equivalent relation, i.e., specific/general concepts. This experiment also investigates the quality of the classification approach, which is used to define the validation strategies, hence, estimate the number of validation tasks (i.e., number of users). In other words, the experiment investigates if the estimated difficulties of the mapping selection tasks confirms the observations concluded from the study in (Abu Helou et al., 2016).

We evaluate the performance of ICLM considering two different configurations; based on the number of validation tasks assigned to each validation difficulty level:

- *BRAVE strategy* (V^L=0, V^M=1, V^H=3), the Low-difficulty tasks will be included into the final alignment without validation; and
- *CAUTIOUS strategy* (V^L=1, V^M=3, V^H=5), every task will be validated by some users.

We evaluate the performance of the alignments found with every configuration against a gold standard. We use the well-know performance measures of Precision, Recall, and F_1-measure, to quantify the performance of the alignments. We compare the two configurations with an alignment automatically obtained using the configuration $TSM + LSOA$ (Abu Helou and Palmonari, 2015), i.e, without validation (V^L=0, V^M=0, V^H=0).

Next, We describe the experimental settings. the gold standard and the users involved in the validation tasks. Further, We describe the validation tasks (steps that users follows). Finally, We report the main results of the experiment. In what follows, we consider the scenario of mapping Arabic concepts to concepts in the English WordNet.

In this experiment six bilingual speakers, from different background: geography, computer science, law, medicine, management, and engineering are asked to link a set of Arabic concepts, taken from the Arabic wordnet (ArWN) (Rodríguez et al., 2008), to concepts in the English WordNet by using ICLM. Users are undergraduate students (2), postgraduate students (2), and doctorates (2). These users are knowledgeable about ICLM and its goals. For each source concept ICLM retrieves the set of candidate matches, which are ranked based on their similarity with the source concept.

Table 3: Sample dataset: distribution by task's difficulty.

	Validation strategy			Total
	Low difficulty	Mid difficulty	High difficulty	
Sample (#concepts)	99	99	52	250
Sample (%)	39.6	39.6	20.8	100.0

Table 4: Sample dataset: distribution by synonym words.

#Synonym words	1	2	3	4	5	6	7	9	Total
Sample (#concepts)	101	72	47	18	6	2	3	1	250

Table 5: Sample dataset: distribution by word type.

Word Type	noun	verb	adjective	adverb	Total
ArWN(%)	68.8	24.3	5.9	1.0	100.0
Sample (#concepts)	166	62	19	3	250
Sample(%)	66.4	24.8	7.6	1.2	100.0

Dataset and Sampling Criteria. We randomly selected 250 concepts from the Arabic wordnet, such that certain condition are satisfied. The concepts are selected to reflect a uniform distribution (w.r.t the gold standard, see Table 2) of concepts category (described in Table 1) as well as tasks difficulty. The following factors are considered while selecting the sample concepts: decidable vs undecidable mappings, the number of synonym words in a source concept, the type (part of speech) of concepts lexicalization, the size of the top-ranked matches in the undecidable mappings, and the position of concepts in the semantic hierarchy (concepts' specialization); i.e., the position that synsets occupy in the semantic hierarchies; such that domain-specific concepts are positioned at lower positions, e.g., synsets that occur as leaf nodes in the semantic hierarchies, where as synsets at top positions express more the general concepts. Tables 2, 3, 4, 5, 6, and 7 report these details.

The validation tasks are processed as follows: After registration, a user can access and start validating the matches. The following instructions (guidelines) are provided to the users:

- *register* to the system and *login*;

Table 8: Performance results: different validation configurations.

	Configuration								
	TSM+LSOA $V^L = V^M = V^H = 0$			BRAVE validation $V^L = 0, V^M = 1, V^H = 3$			CAUTIOUS validation $V^L = 1, V^M = 3, V^H = 5$		
Relation	R	P	F1	R	P	F1	R	P	F1
Equivalent	0.50	0.50	0.50	0.596	0.667	0.630	0.684	0.718	0.701
Equivalent, Specific, General	-	-	-	0.672	0.676	0.674	0.796	0.734	0.764
# Required validation	-			[203-255]			[453-650]		
# Preformed validation	-			233			556		
# Avg. time/validation (sec)	-			97			89		
# Equivalent relation	-			149			171		
# Specific relation	-			11			16		
# General relation	-			8			12		

Table 6: Sample dataset: distribution by (noun) concepts specialization.

Position in the hierarchy	[1-3]	[4-6]	[7-9]	[10-12]	[13-15]	Total
Sample (#concepts)	2	47	88	26	2	166

Table 7: Sample dataset: distribution by TopSet cardinality.

Cardinality of TopSet	[4-10]	[11-20]	[21-40]	Total
Sample (#concepts)	93	24	8	125

- select the respective language (*Arabic*) of the source concept list;
- select the *Full List* of candidate matches;
- select one of the source concepts from the Arabic concept (*Source Concepts*);
- evaluate the list of candidate matches (*Top Candidate Matches*);
- if the lexicalization (synonyms) of a candidate matches is not sufficient to validate the mapping, click on the candidate match for getting more details. In the *Matches Info* side one can find more useful details, which includes definitions, examples, and neighbor (parent and sibling) concepts. These information are navigable to the online English WordNet. Similarly, an online Arabic glossary (Al-maany glossary website) is also accessible and linked to each source synonym word (see Figure 3). Use the full-text search if a correct candidate match does not appear in the top positions;
- once identified the potentially correct candidate match, choose one of the following relationships:
  - General ($\Uparrow$): the candidate concept is more generic with respect to the source concept;
  - Equivalent ($\Leftrightarrow$): the candidate concept is equivalent to the source concept;
  - Specific ($\Downarrow$): the candidate concept is more specific with respect to the source concept;
- select another concept from the source list until all the concept have been evaluated.

5.1 Results and Discussion

Table 8 reports the performance measures for the three configurations. Precision (P) measures how many selected relation are correct w.r.t the gold standard. Recall (R) measures how many correct relations are selected w.r.t the gold standard. F_1-measure is the harmonic mean of the two measures. The first row reports the performance of selecting the equivalent relations, while the second row reports if also specific or general relations are also correctly selected. The third row reports the required number of validations: the lower bound refers to the minimum number of validations, which happen if a consensus agreement occurs for each source concept; whereas the upper bound refers to the maximum number of validations when no agreement achieved. The fourth row reports the number of validations performed by the users. The average elapsed time that users spent to validate a mapping is reported in the last row. Observe that, the performance without validation, in the first column, is 50%, since 50% of the sample dataset (Table 2) are decidable mappings, i.e., the candidate matches include the correct match that is ranked as TopOne.

Table 8 reports that the average elapsed time in the CAUTIOUS validation is less than the time in the BRAVE validation; one reason might be due to the increase of users awareness of the system.

The last three rows in Table 8 report the number of relations that users define through the selection tasks. The defined relationships are split as follows. In the BRAVE validation; 11 of type specific relation, 8 of type general relation, and 149 of type equivalent relation; in the CAUTIOUS validation: 16 of type specific relation, 12 of type general relation, and 171 of type equivalent relation. Based on the minimum consensus agreement approach users effort is reduced by

5.8% and 15.4% in the BRAVE and CAUTIOUS validations restrictively, w.r.t the maximum number of the required validations.s

Moreover, an important observation is that, users have not reached an agreement in the High-difficulty validation in most cases. This is due to the fact that the available evidence, even for the users, are not sufficient to decide and select the correct relation. If definitions or examples (sense tagged sentences) are available for the source concepts, i.e., any further contextual knowledge, it would be easier for the users to select the correct relation. For instance, in most of the High-difficulty validations users accessed the online glossary aiming to find more evidence, however, the glossary provides all the possible definitions (senses) of the word without disambiguating its sense. While information provided about the candidate matches (*Matches Info*) seems to be sufficient for the users, few of them accessed the online WordNet in order to navigate the wordnet hierarchic. In fact, this confirms the usefulness of the classification method defined (Abu Helou et al., 2016), and the efficiency in estimating the difficulty of the mapping selection tasks based on the available evidence.

6 CONCLUSIONS

The paper presented a semi-automatic cross-lingual matching system using multi-user feedback scenario, called Interactive Cross-Lingual Mapping (ICLM). ICLM is a Web application that supports users with quality mappings by leveraging translation evidence and lexical characteristics using a lexical based disambiguation algorithm. ICLM reduces the users effort by distributing the mapping selection tasks to a different number of users based on an estimated difficulty of these mappings, and accordingly collects users feedback in more efficient way, in contrast to pure crowdsourcing models where tasks are equally assigned to a fixed number of users. A user study is conducted to evaluate ICLM's strategies in estimating and distributing the validation tasks. The experimental results provide evidence that the estimated difficulties to a large extent are precise, and the classification method used to classify these task is useful.

As a future direction, we plan to use and adapt the ICLM approach described here to support matching of lexical resources in the context of the EW-Shopp [5] (Supporting Event and Weather-based Data Analytics and Marketing along the Shopper Journey) and EuBusinessGraph [6] (Enabling the European Business Graph for Innovative Data Products and Services) H2020 EU projects. In particular, ICLM can be helpful to support alignment of resources such as product descriptions and categories, or business classification systems, published in different languages in Europe.

In the future, we also plan to investigate further strategies to distribute the selection tasks over users. For instance, we would like to investigate an active learning model presented in (Cruz et al., 2014). Another interesting direction would be to consider more languages and incorporate more users. In addition, to learn from users behavior in order to reconfigure the difficulty estimation is another interesting direction to explore. Moreover, an in-depth analyze w.r.t each concept category should be also considered.

[5] www.ew-shopp.eu

[6] http://eubusinessgraph.eu/

ACKNOWLEDGEMENTS

The work presented in this paper has been partially supported by EU projects funded under the H2020 research and innovation programme: EW-Shopp, Grant n. 732590, and EuBusinessGraph, Grant n. 732003.

REFERENCES

Abu Helou, M. and Palmonari, M. (2015). Cross-lingual lexical matching with word translation and local similarity optimization. In *Proceedings of the 10th International Conference on Semantic Systems, SEMANTiCS 2015, Vienna, Austria, September*.

Abu Helou, M., Palmonari, M., and Jarrar, M. (2016). Effectiveness of automatic translations for cross-lingual ontology mapping. *J. Artif. Intell. Res. (JAIR) Special Track on Cross-language Algorithms and Applications*, 55:165–208.

Cruz, I., Loprete, F., Palmonari, M., Stroe, C., and Taheri, A. (2014). Pay-as-you-go multi-user feedback model for ontology matching. In Janowicz, K., Schlobach, S., Lambrix, P., and Hyvnen, E., editors, *Knowledge Engineering and Knowledge Management*, volume 8876 of *Lecture Notes in Computer Science*, pages 80–96. Springer International Publishing.

Cruz, I. F., Antonelli, F. P., and Stroe, C. (2009). Agreementmaker: Efficient matching for large real-world schemas and ontologies. *PVLDB*.

Cruz, I. F., Palmonari, M., Loprete, F., Stroe, C., and Taheri, A. (2016). Quality-based model for effective and robust multi-user pay-as-you-go ontology matching. *Semantic Web*, 7(4):463–479.

Cruz, I. F., Stroe, C., and Palmonari, M. (2012). Interactive user feedback in ontology matching using signature vectors. In Kementsietsidis, A. and Salles, M. A. V., editors, *ICDE*, pages 1321–1324. IEEE Computer Society.

de Melo, G. and Weikum, G. (2012). Constructing and utilizing wordnets using statistical methods. *Language Resources and Evaluation*, 46(2):287–311.

Demartini, G., Difallah, D. E., and Cudr-Mauroux, P. (2012). Zencrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In Mille, A., Gandon, F. L., Misselis, J., Rabinovich, M., and Staab, S., editors, *WWW*, pages 469–478. ACM.

Dragisic, Z., Ivanova, V., Lambrix, P., Faria, D., Jimenez-Ruiz, E., and Pesquita, C. (2016). User validation in ontology alignment. In *Proceedings of the International Semantic Web Conference*, volume 9981 of *LNCS*.

Fellbaum, C., editor (1998). *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA ; London.

Garcia, J., Montiel-Ponsoda, E., Cimiano, P., Gómez-Pérez, A., Buitelaar, P., and McCrae, J. (2012). Challenges for the multilingual web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 11(0).

Hovy, E., Navigli, R., and Ponzetto, S. P. (2012). Collaboratively built semi-structured content and artificial intelligence: The story so far. *Artificial Intelligence*.

Jimnez-Ruiz, E., Grau, B. C., Zhou, Y., and Horrocks, I. (2012). Large-scale interactive ontology matching: Algorithms and implementation. In Raedt, L. D., Bessire, C., Dubois, D., Doherty, P., Frasconi, P., Heintz, F., and Lucas, P. J. F., editors, *ECAI*, volume 242 of *Frontiers in Artificial Intelligence and Applications*, pages 444–449. IOS Press.

Narducci, F., Palmonari, M., and Semeraro, G. (2013). Cross-language semantic matching for discovering links to e-gov services in the lod cloud. In Vlker, J., Paulheim, H., Lehmann, J., Niepert, M., and Sack, H., editors, *KNOW@LOD*, volume 992 of *CEUR Workshop Proceedings*, pages 21–32. CEUR-WS.org.

Narducci, F., Palmonari, M., and Semeraro, G. (2017). Cross-lingual link discovery with TR-ESA. *Inf. Sci.*, 394:68–87.

Navigli, R. and Ponzetto, S. P. (2012). Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.*, 193:217–250.

Noy, N. F. and Musen, M. A. (2003). The PROMPT suite: interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies*, 59(6):983–1024.

Pazienza, M. and Stellato, A. (2006). An open and scalable framework for enriching ontologies with natural language content. In Ali, M. and Dapoigny, R., editors, *Advances in Applied Artificial Intelligence*, volume 4031 of *Lecture Notes in Computer Science*, pages 990–999. Springer Berlin Heidelberg.

Pianta, E., Bentivogli, L., and Girardi, C. (2002). Multiwordnet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*.

Rodríguez, H., Farwell, D., Farreres, J., Bertran, M., Mart, M. A., Black, W., Elkateb, S., Kirk, J., Vossen, P., and Fellbaum, C. (2008). Arabic wordnet: Current state and future extensions. In *Proceedings of the Forth International Conference on Global WordNet*.

Sarasua, C., Simperl, E., and Noy, N. F. (2012). Crowdmap: Crowdsourcing ontology alignment with microtasks. In *Proceedings of the 11th International Semantic Web Conference (ISWC2012)*.

Shi, F., Li, J., Tang, J., Xie, G. T., and Li, H. (2009). Actively learning ontology matching via user interaction. In Bernstein, A., Karger, D. R., Heath, T., Feigenbaum, L., Maynard, D., Motta, E., and Thirunarayan, K., editors, *International Semantic Web Conference*, volume 5823 of *Lecture Notes in Computer Science*, pages 585–600. Springer.

Shvaiko, P. and Euzenat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Trans. Knowl. Data Eng.*, 25(1):158–176.

Spohr, D., Hollink, L., and Cimiano, P. (2011). A machine learning approach to multilingual and cross-lingual ontology matching. In *International Semantic Web Conference (1)*, pages 665–680.

Trojahn, C., Fu, B., Zamazal, O., and Ritze, D. (2014). *State-of-the-art in Multilingual and Cross-Lingual Ontology Matching*. Springer.

Zhang, Z. (2014). Towards efficient and effective semantic table interpretation. In *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I*, pages 487–502.