Overview of the NTCIR-9 Crosslink Task: Cross-lingual Link Discovery

Ling-Xiang Tang[1] , Shlomo Geva[1] , Andrew Trotman[2] , Yue Xu[1] and Kelly Y. Itakura[1] [1]Faculty of Science and Technology, Queensland University of Technology [2]Department of Computer Science , University of Otago

ABSTRACT

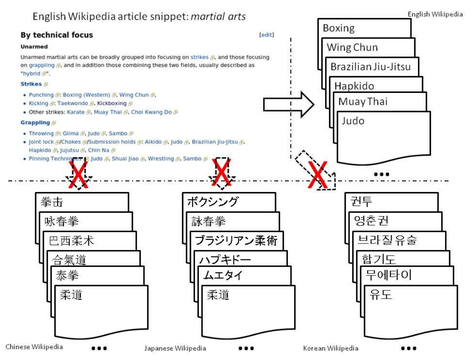
This paper presents an overview of NTCIR-9 Cross-lingual Link Discovery (Crosslink) task. The overview includes: the motivation of cross-lingual link discovery; the Crosslink task definition; the run submission specification; the assessment and evaluation framework; the evaluation metrics; and the evaluation results of submitted runs.

1. WHAT IS CROSS-LINGUAL LINK DISCOVERY

Cross-lingual link discovery (CLLD) is a way of automatically finding potential links between documents in different languages. The goal of this task is to create a reusable resource for evaluating automated CLLD approaches. The results of this research can be used in building and refining systems for automated link discovery. The task is focused on linking between English source documents and Chinese, Korean, and Japanese target documents.

2. WHY CLLD?

Wikipedia is an online multilingual encyclopaedia that contains a very large number of articles covering most written languages and so it includes extensive hypertext links between documents of same language for easy reading and referencing. However, the pages in different languages are rarely linked except for the cross-lingual link between pages about the same subject. This could pose serious difficulties to users who try to seek information or knowledge from different lingual sources. The Figure on the right shows a snippet of *Martial Art* Wikipedia article in which anchors are only linked to related English articles about different types of martial arts; direct links to other related Chinese/Japanese/Korean articles do not exist in Wikipedia.



3. CLLD TASK DEFINITION

CLLD TASK

To submit a run for a given task, participants are required to choose the most suitable anchors in English topic documents, and for each anchor identify the most relevant documents in the target language corpus.

SUBTASKS

- English to Chinese CLLD (E2C)
- English to Japanese CLLD (E2J)
- English to Korean CLLD (E2K)

TOPICS

A set of 25 articles are randomly chosen from the English Wikipedia and used as test topics. All test topics are prepared in a form of XML file without *link* tags, which means the previously existing links in topics are removed.

WIKIPEDIA TEST COLLECTIONS

Language	# doc	Size	Dump Date
Chinese	318736	2.7G	27/06/2010
Japanese	716,088	6.1G	24/06/2010
Korean	201596	1.2G	28/06/2010

4. ASSESSMENT AND EVALUTION

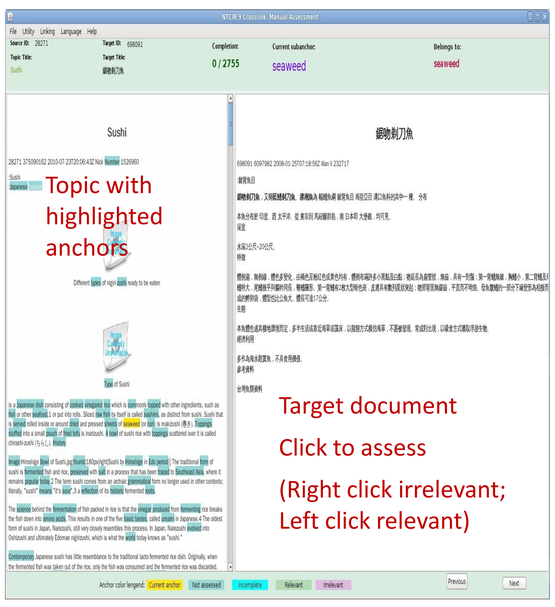
ASSESSMENT

- Wikipedia ground-truth (existing links)
- Manual Assessment

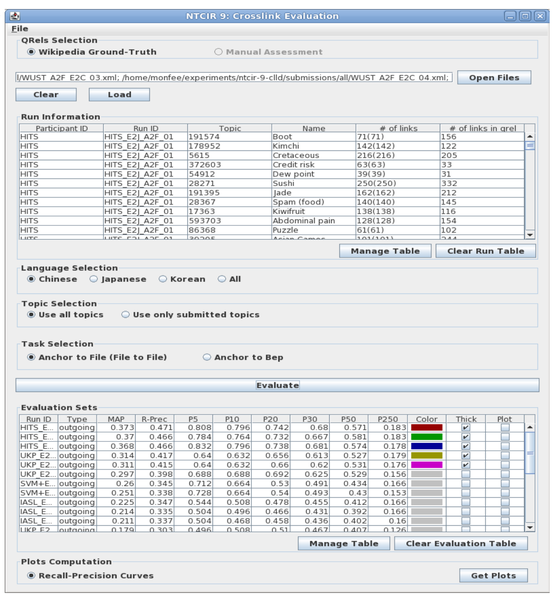
EVALUATION

- File-to-File evaluation
- Anchor-to-File evaluation

ASSESSMENT TOOL



EVALUATION TOOL



5. EVALUATION METRICS

PRECISION AND RECALL IN F2F EVALUATION

$$\textbf{Precision}_{f2f} = \frac{\textbf{number of correct links}}{\textbf{number of identified links}}$$

$$\textbf{Recall} = \frac{\textbf{number of correct links}}{\textbf{number of links in qrels}}$$

Where *n* is the number of identified anchors; *N* is the number of anchors in *qrel*; *k* is the number of returned targets for anchor *i;* and k_i is the number of targets recommended for this anchor.

PRECISION AND RECALL IN A2F EVALUATION

$$f_{anchor}(i) = \begin{cases} 1, & \text{if relevant with} \geq 1 \text{ relevant targets} \\ 0, & \text{otherwise} \end{cases}$$

$$f_{link}(j) = \begin{cases} 1, & \text{if relevant} \\ 0, & \text{otherwise} \end{cases}$$

$$\textbf{Precision}_{a2f} = \left(\sum_{i=1}^{n}(f_{anchor}(i)) \times \frac{\sum_{j=1}^{k} f_{link}(j)}{k_i}\right)/n$$

$$\textbf{Recall}_{a2f} = \left(\sum_{i=1}^{n}(f_{anchor}(i)) \times \frac{\sum_{j=1}^{k} f_{link}(j)}{k_i}\right)/N$$

SYSTEM EVALUATION METRICS: *MAP, R-PREC, P@N*

$$\textbf{MAP} = \left(\sum_{t=1}^{n} \frac{\sum_{k=1}^{m} p_{kt}}{m}\right)/n$$

where n is the number of topics (source articles used in evaluation); m is the number of identified items (articles for F2F or anchors in A2F); and P_{kt} is the precision of the top K items for topic t.

$$\textbf{R-Prec} = \sum_{t=1}^{n} P_t @ R \,/n$$

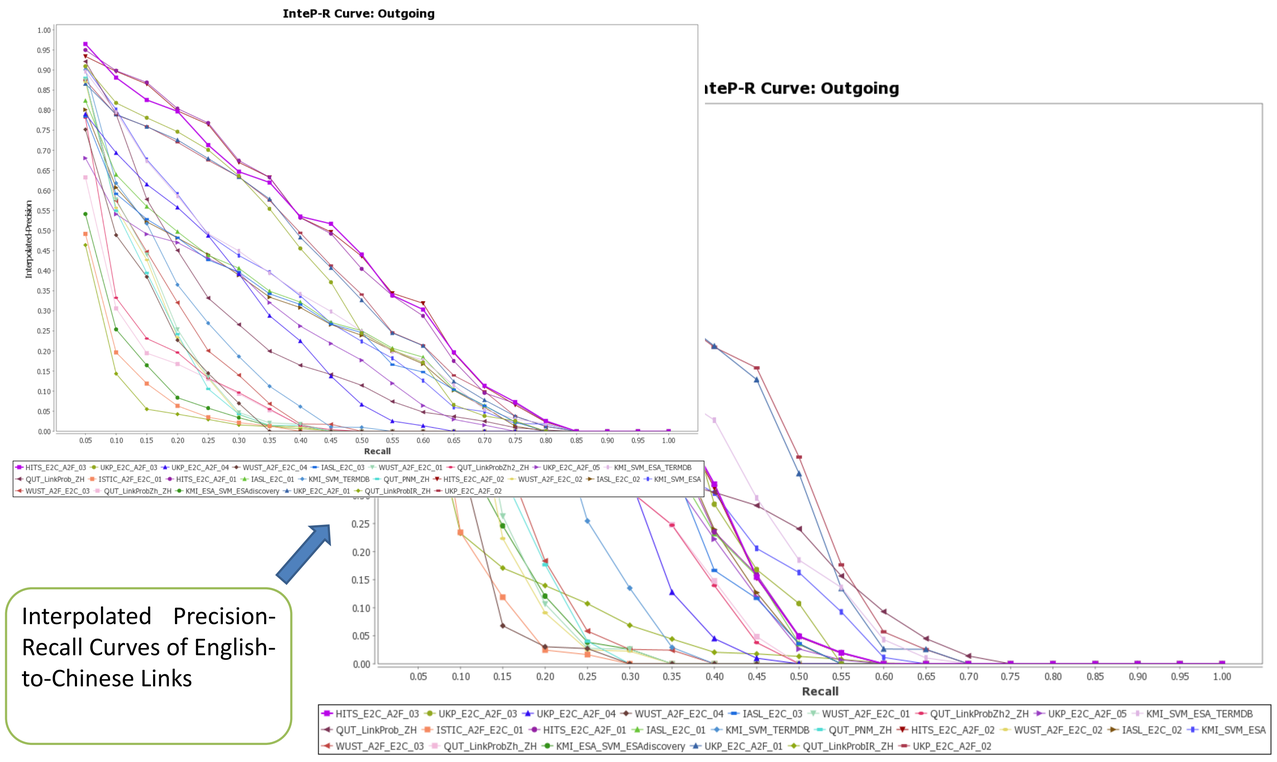
where n is the number of topics; and $P_t@R$ is the precision at R where R is the number of unique items in the *qrel*s of topic t.

Precision-at-N is computed using the average precision for all topics (source articles) at a pre-defined position N in the results list. Values of N were chosen as: 5, 10, 20, 30, 50, and 250.

CONCLUSION

- The evaluations in both file-to-file and anchor-to-file levels show promising results of participating teams even this is the first year of this task. Particularly, team HITS and UKP achieved very high scores of different evaluation measures (*MAP, R-Prec, Precision-at-N*) in three language subtasks.
- With the standard test collections and evaluation data set developed for this task, systems or applications for realising CLLD can be built or further refined to provide a better automated cross-lingual linking for knowledge management and sharing. With the possible future deployment of techniques and approaches used by the participants, we hope user experience in viewing or editing document in different languages could be enhanced; language is then no longer a barrier for knowledge discovery.

7. EVAULATION RESULTS



Interpolated Precision-Recall Curves of English-to-Chinese Links

THE TOP THREE TEAMS OF EACH MEASURE (*MAP, R-PREC, P@N)*

F2F Wiki-GT EVALUATION	F2F MANUL EVALUATION	F2F MANUL EVALUATION
English-2-Chinese ***MAP:*** HITS, UKP, KMI ***R-Prec:*** HITS, UKP, KMI ***Precision-at-5:*** HITS, QUT, KMI	*English-2-Chinese* ***MAP:*** UKP, KMI, HITS ***R-Prec:*** UKP, KMI, HITS ***Precision-at-5:*** QUT, HITS, KMI	*English-2-Chinese* ***MAP:*** UKP, QUT, HITS ***R-Prec:*** UKP, QUT, KMI ***Precision-at-5:*** KMI, QUT, UKP
English-2-Japanse ***MAP:*** HITS, UKP, QUT ***R-Prec:*** HITS, UKP, QUT ***Precision-at-5:*** HITS, UKP, QUT	*English-2-Japanse* ***MAP:*** HITS, UKP, QUT ***R-Prec:*** HITS, UKP, QUT ***Precision-at-5:*** HITS, UKP, QUT	*English-2-Japanse* ***MAP:*** HITS, UKP, QUT ***R-Prec:*** HITS, UKP, QUT ***Precision-at-5:*** HITS, UKP, QUT
English-2-Korean ***MAP:*** HITS, DUIIS, UKP ***R-Prec:*** HITS, DUIIS, UKP ***Precision-at-5:*** HITS, DUIIS, KSLP	*English-2-Korean* ***MAP:*** UKP, DUIIS, HITS ***R-Prec:*** UKP, DUIIS, HITS ***Precision-at-5:*** QUT, IISR, HITS	*English-2-Korean* ***MAP:*** UKP, HITS, KSLP ***R-Prec:*** UKP, HITS, KSLP ***Precision-at-5:*** HITS, KSLP, UKP

The teams with the highest number of unique relevant links found with the Wikipedia ground-truth are: ***UKP*** for E2C, ***QUT*** for E2J, ***HITS*** for E2K. The teams with the highest number of unique relevant links found with the manual assessment results are: ***QUT*** for E2C, ***HITS*** for E2J, ***KSLP*** for E2K

6. SUBMISSIONS

GROUP	ORGANISATION
DUIIS	Daegu University
HITS	Heidelberg Institute for Theoretical Studies
IISR	Yuan Ze University
ISTIC	Institute of Scientific and Technical Information of China
KMI	The Open University
kslab_nut	Nagaoka University of Technology
KSLP	Kyungsung University
nthuisa	Academia Sinica
QUT	Queensland University of Technology
UKP	TU Darmstadt
WUST	Wuhan University of Science and Technology

Group	En-2-Zh	En-2-Ja	En-2-Ko
DUIIS	0	0	2
HITS	3	3	3
IISR	0	0	5
ISTIC	1	0	0
KMI	4	0	0
kslab_nut	0	1	0
KSLP	0	0	5
nthuisa	3	0	0
QUT	5	2	1
UKP	5	5	5
WUST	4	0	0
Sub-total	25	11	21
Total	57		