NTCIR9-GeoTime Overview - Evaluating Geographic and Temporal Search: Round 2

Fredric Gey†, Ray Larson†, Jorge Machado ‡, Masaharu Yoshioka*
†University of California, Berkeley USA
‡INESC-ID, National Institute of Electroniques and Computer Systems, Lisbon, PORTUGAL
* Hokkaido University, Sapporo, JAPAN
gey@berkeley.edu , ray@ischool.berkeley.edu ,
jorge.r.machado@ist.utl.pt , yoshioka@ist.hokudai.ac.jp

ABSTRACT

GeoTime for the NTCIR Workshop 9 is the second evaluation of Geographic and Temporal Information Retrieval called “NTCIR GeoTime”. The focus of this task is on search with Geographic and Temporal constraints. This overview describes the data collections (Japanese and English news stories), topic development, assessment results and lessons learned from this second NTCIR GeoTime task, which combines GIR with time-based search to find specific events in a multilingual collection. Six teams submitted Japanese runs and nine teams submitted English runs. Three teams participated in both Japanese and English.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval—retrieval models, search process**. General Terms:** Experimentation, Performance, Measurement **Keywords:** Crosslingual Information Retrieval; Geotemporal Search, Geographic Information Retrieval, IR evaluation

1. INTRODUCTION

Cultural Geographic search is quite prevalent in many modern search venues. A great number of documents (web, news, and scientific) have a geographic focus. Geographic search allows for a unique user interface, the interactive map, which can be utilized not only to narrow the user’s focus by geography, but also to highlight interesting events. There have been over six workshops [9] on Geographic Information Retrieval (GIR) held in association with SIGIR, CIKM, ECDL or other conferences as well as workshops and conference tracks on location-based search, there has also been 4 years of evaluation of GIR within CLEF (the GeoCLEF track). But, until this task at NTCIR, Asian language geographic search had never been specifically evaluated, even though about half of the NTCIR-6 Cross-Language topics had a geographic component (usually a restriction to a particular country).

Geographic information retrieval is concerned with the retrieval of thematically and geographically relevant information resources in response to a query of the form {<*theme* or *topic*, *spatial relationship*, *location*>}, e.g. ``Temples within 5 km. of Tokyo". [6]. Systems that support GIR, such as geographic digital libraries, and location-aware web search engines, are based on a collection of georeferenced information resources and methods to spatially search these resources with geographic location as a key. Information resources are considered geo-referenced if they are spatially indexed by one or more regions on the surface of the Earth, where the specific locations of these regions are encoded either directly as spatial coordinates, i.e. geometrically, or indirectly by place name [4]. However, in order for place names to support a spatial approach to GIR, they must be associated with a model of geographic space.

The temporal aspects of search have been largely ignored in the IR community, but not in the GIS and information processing communities. There has been a special issue of ACM TALIP on 'Temporal Information Processing' [8], as well as at least two workshops on "Temporal and Spatial Information Processing". The NTCIR-GeoTime organizers wanted to utilize and incorporate past research on this aspect as part of the evaluation. Only recently, with GeoTime for NTCIR-8 [2] and with a paper in the most recent GIR workshop at the University of Zurich [12] has the combination of Geographic and Temporal search been addressed.

2. DATA

For GeoTime in NTCIR-8, two news story collections were used, one Japanese and one English. The Japanese collection consisted of Mainichi newspapers for 2002-2005, while the English collection, consisted of New York Times stories also for 2002-2005. For GeoTime in NTCIR-9 we wished to add a Korean collection, but the intellectual property rights could not be negotiated in a timely manner. In addition, a Korean co-organizer could not be found who had the time to participate.

So, as with NTCIR-8 GeoTime, the collections were both English and Japanese. However we were able to significantly expand both the size and time coverage of news collections for these languages by adding documents from earlier NTCIR workshops. In particular the Mainichi Japanese News collection was expanded by adding stories for the period 1998-2001. Because NY Times documents were unavailable for 1998-2001, to cover the same time period in English, the Xinhua Chinese news service English subsection and the English documents from the Korea Times were added and from the English edition of Mainichi. Details about these collections are found in Table 1.

Collection	Language	Time Period	# Documents
Mainichi	J	1998-2001	419,759
Mainichi	J	2002-2005	377,941
Korea Times	E	1998-2001	50,129
Mainichi	E	1998-2001	24,878
NY Times	E	2002-2005	315,417
Xinhua	E	1998-2001	406,792

Table 1: Collections used for NTCIR9-GeoTime

Users of the English collections had to pay a fee of $50US to the Linguistic Data Consortium to prepare and mail the DVD with the NY Times collection or download the Korea Times and Xinhua collections.

In GeoTime for NTCIR-8 we discovered gaps in the NYT collection for Jan 2003-July 2004. Details about this can be found in [2]. Since in topic development we wished to create topics which had relevant documents in both collections, we had to shy away from events which happened in 2003-June 2004.

3. TOPIC DEVELOPMENT

Similar to GeoTime for NTCIR-8, for this round of GeoTime, we invited participating groups to submit possible topics. The guidelines were to propose questions with time and space aspects. The ground truth evidence was to be provided by Wikipedia articles which specified both time and place. As with the NTCIR8 GeoTime, most topics could be found in the annual notable events in Wikipedia, For example, topic GeoTime-0035 (*In an African country an oil pipeline explosion killed more than 700 people – when and where did this occur*) is found in e.g. http://en.wikipedia.org/wiki/1998, pointing the article: http://en.wikipedia.org/wiki/1998_Jesse_pipeline_explosion. From a geographic point of view, this makes our evaluation seem to resemble GikiCLEF [7] the CLEF 2009 track which asked questions against a multilingual subset of Wikipedia.

As with GeoTime for NTCIR-8, organizer Ray Larson indexed both the English and Japanese collections using his Cheshire system, and provided a search engine for testing topics against the collection. In addition, organizer Masaharu Yoshioka also created a Japanese search engine for the Japanese collection. These engines (password protected) were made available to participating teams. With this process the organizers created another 25 topics in either English or Japanese which were translated into the other language. Each of the 25 topics was vetted to hit at least two relevant documents in both languages.

Of some note is that we inserted slight variations of a single topic in topics 35 (above), 36, and 37. Topic 36 (*When and where have there been pipeline explosions in an African country with more than 5 fatalities?)* was a variation that required an open-ended list to answer, while topic 37 (*What fatal accident occurred near geographical coordinates 5° 52′ 12″ N 5° 45′ 00″ E / 5.870 ° N 5.750° E / 5.870; 5.750), which killed hundreds of people, and when did it occur?*) required reverse geocoding for effective retrieval. More information can be found in our EVIA paper [3].

More discussion and evaluation of topic difficulty will follow the presentation of results.

4. PARTICIPATION

While twenty groups signed up to participate in NTCIR-GeoTime, only twelve groups actually submitted runs, of whom three were organizers of the evaluation.

Japanese runs were submitted by the following six groups

BRKLY	University of California, Berkeley*
HU-KB	Hokkaido University, Japan*
KOLIS	Keio University, Library Science
NAK	Keio University, Science and Technology
OKSAT	Osaka Kyoiku University, Japan
RMIT	Royal Melbourne Institute of Technology

English runs were submitted by the following nine groups:

Team Name	Organization
BRKLY	University of California, Berkeley, USA*
GETUA	University of Alicante, Spain
INESC	National Institute of Electroniques & Computer Systems, Lisbon, Portugal *
IRNLP	Korea Advanced Inst. for Science & Technology
OKSAT	Osaka Kyoiku University, Japan
RMIT	Royal Melbourne Institute of Technology
SINAI	University of Jaén, Spain
SJTUB	Shanghai Jiao Tong University, China
UIOWA	University of Iowa, USA

*Organizer group

Each group was allowed to submit up to 5 runs per source-target language combination. While we encouraged the submission of bilingual runs, only BRKLY submitted such runs for JP→EN, and BRKLY and the KOLIS Japanese group submitted EN→JP runs. The following table summarizes the number of runs submitted by each group:

Team	JA→JA	EN→EN	EN→JA	JA→EN
BRKLY	2	2	2	2
GETUA		5		
HU-KB	5			
INESC		5		
IRNLP		4		
KOLIS	5		5	
NAK	2			
OKSAT	5	5		
RMIT	5	2		
SINAI		4		

SJTUB		4		
UIOWA		4		

5. EVALUATION

Relevance judging was done in a traditional manner on a pool of the top 100 documents retrieved from all runs with duplicates removed. Relevance assessment for Japanese was undertaken by teams at NII and Hokkaido University using the SEPIA system utilized for ACLIA and IR4QA in NTCIR-8. For English assessment, the third author had developed a system at Technical University of Lisbon in Portugal and English assessment was done worldwide with assessors in Australia, Japan, Portugal, Spain and USA. This was the second time for this process for NTCIR-GeoTime. English participating teams were assigned to assess an approximately equal number of documents (as close to the mean number of pool documents over the 25 topics) For Japanese GeoTime, 15,795 documents were examined and judged. For the English GeoTime, 19,966 were examined and judged. Judgment was graded in that a document could be assessed as "fully relevant" if it contained text which answered both the "when" and "where" aspects of the topic. The document was assessed as 'partially relevant – where' if it answered the geographic aspect of the topic and 'partially relevant – when' if it answered the temporal aspect of the topic. In order to utilize existing evaluation software, the three fully and partially relevant categories were aggregated into a single category upon which the following result tables are based. We hope to have a more detailed analysis separating out the categories in the final paper.

6. APPROACHES

The approaches taken by participating teams in GeoTime for NTCIR-9 were, overall, more sophisticated that for the previous GeoTime in NTCIR-8 [2]. There was significant reliance on named entity recognition for geography as well as time tagging. Several teams used geographic filtering (INESC) or other Boolean query construction (HU-KB, UIOWA) to take advantage of the entity recognition. HU-KB from Hokkaido University transformed from day-of-week expressions to absolute dates by matching with document date. The SJTUBCMI team from Shanghai Jiao Tong University used a passage-based learning-to-rank approach. The KOLIS Group from Keio University attempted to improve on their rather unique re-ranking algorithm of NTCIR8-GeoTime where they counted the number of geographic and temporal expressions and re-ranked according to this count. Sophisticated NLP was also applied by Korea Advanced Institute for Science and Technology (NLPIR) by doing Semantic Role Labeling in combination with multiple models and rank aggregation. A number of teams made more extensive use of external resources such as Wikipedia, DBpedia, Geonames, Alexandria Digital Library and Yahoo! PlaceMaker. In NTCIR-8, such resources were only used by a few groups.

The best-performing team from Osaka Kyoiku University (OKSAT) used external resources such as Wikipedia and Google Maps to construct queries from topics by having team members extract time and place from Wikipedia documents and manually inserting them into the text of the query. In particular for topic 37, they searched Google Maps for the latitude/longitude location and manually extracted place names found in that neighborhood to add to the final query. In a sense we could say that the team constructed queries which included the essence of the answer to the question posed by the topic. Their results using this approach substantially outperformed other teams' runs. In a sense, their runs provide a goal post to be attained by fully automatic methods. In addition, the OKSAT runs probably retrieved many relevant documents which might not otherwise have been included into the assessment pool. In the results below, their manual runs which included human effort in query construction are marked as such.

7. RESULTS

7.1 English Results

For search against the English collections, the nine groups submitted 37 runs. Table 1a summarizes the results for English sorted by the mean performance over 25 topics showed for three performance measures, Average Precision (AP), Q, and normalized Discounted Cumulative Gain (nDCG). As can be seen from the table, the top performing runs were very close, but performance order differs depending upon metric. The top 10 runs are in identical order for AP and Q; however the order changes substantially when using the nDCG measure. For direct comparison of best results by team, we selected the best team result for description only runs, found in Table 1b.

7.2 Japanese Results

For search against the Mainichi Japanese news collection, eight teams submitted 31 runs whose performance is summarized in Table 2a. Table 2b provides best team performance using topic description only and omitting the narrative.

7.3 Topic Difficulty

We can also make an attempt to assess the difficulty of particular topics for both the English and Japanese collections. Figures 1 and 2 average the three performance measures over all submitted runs and plot this average by topic. The data are sorted by average precision in order to more clearly identify which topics presented the most challenge to successful search.

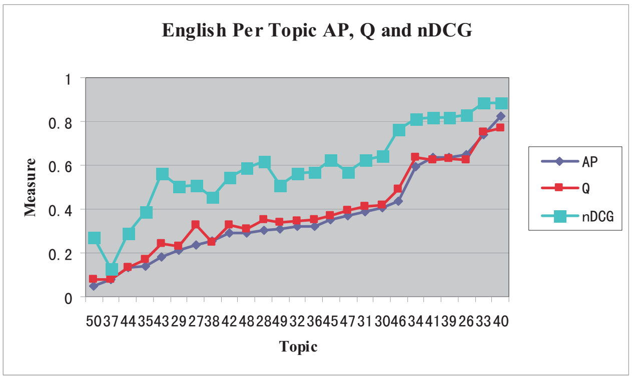

Figure 1: Performance over 37 English runs for 25 topics (pool depth 100), sorted by topic difficulty (AP ascending)

From the point of view of search of the English collections, the four most difficult topics (less than 0.17 overall average precision) seem to be topics 50, 37, 44 and 26.

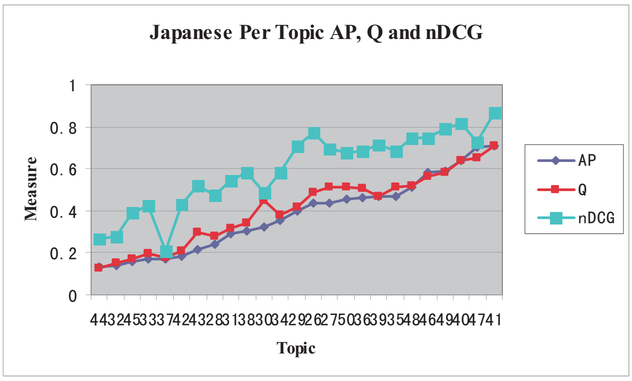

Figure 2: Per-topic AP, Q and nDCG averaged over 31 Japanese runs for 25 topics (pool depth 100), sorted by topic difficulty (AP ascending)

In this round (NTCIR-9 GeoTime) of evaluation, no Japanese topic had less than 0.1 overall average precision while four topics did for GeoTime in NTCIR-8.

7.4 Performance Variability across Topics

Another way to assess performance is to examine individual performance variability across topics. Such performance can be displayed by taking individual topic runs and finding the minimum, median and maximum performance for that topic. These are displayed in Figures 3 (English runs) and 4 (Japanese runs).

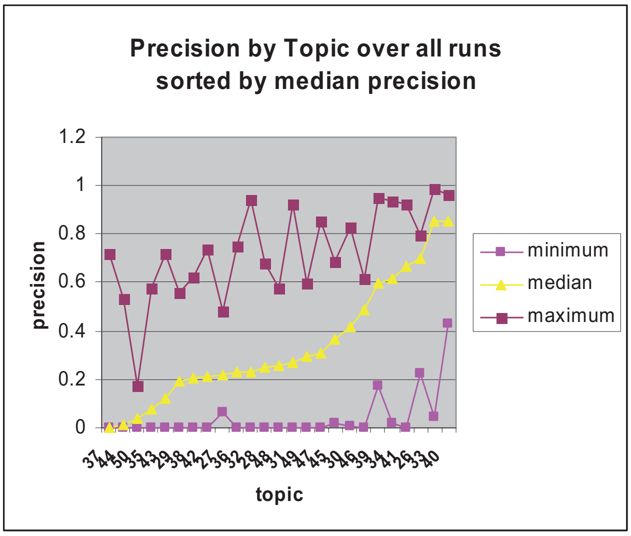

Figure 3: Per-topic AP showing Minimum, Median and Maximum performance for English runs

The interesting feature of this graph is that it clearly shows that for nearly all topics, some group performed substantially better than the median performance for that topic. This is quite different from GeoTime in NTCIR-8 [10] where, for a number of topics, the maximum and median precisions were very close.

The most difficult topics for English seem to be 37 (median 0.0), 44 (median 0.0114), 50 (median 0.0415), and 35 (median 0.0794) all of which had median precision below 0.1. Topic 37 is worth discussing. Topic 37 specified a latitude and longitude and asked what disaster occurred near this location which killed hundreds of people. Only one group (OKSAT) did well on this topic for both English and Japanese – for English their precision was .7207, compared to the median precision of 0.0. The answer was expected to be same as for Topic 35 (Jesse, Nigeria pipeline explosion which killed around 1000 people). However there were other explosions in "nearby" cities (Warri, Nigeria) which had more than 100 fatalities, so we allowed such events to be relevant documents.

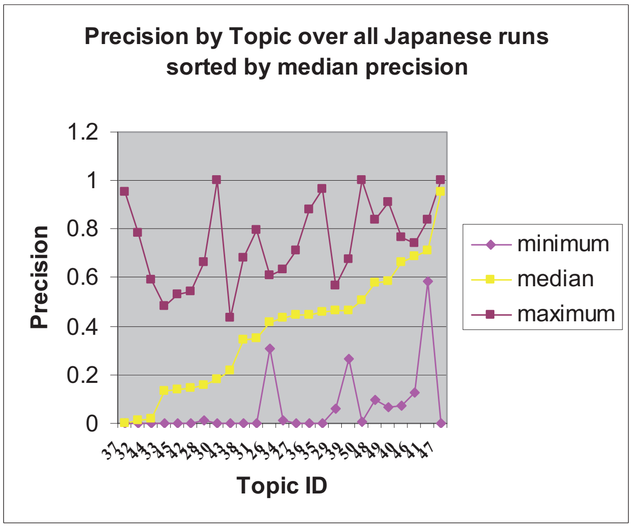

Figure 4: Per-topic AP showing Minimum, Median and Maximum performance for Japanese runs

While for nearly all Japanese topics, at least one group had a minimum precision of near zero for that topic, there was still a wide variability of performance from both minimum to median average precision for a topic, as well as from median precision to maximum precision for a topic. Only in the four topics (40, 46, 41, and 47) with greatest median precision do the median and maximum become close. Where the median and maximum are very close, we can infer that almost all groups had good performance

Table 1a: GeoTime English mean performance for three performance metrics for 37 submitted runs (sorted by descending Mean Average Precision (MAP) over all topics

RunName	MAP	Q	nDCG@10	nDCG@100	nDCG@1000
OKSAT-EN-EN-02-DN*	0.5549	0.5738	0.7379	0.6883	0.7654
OKSAT-EN-EN-03-D*	0.5376	0.5562	0.7311	0.6777	0.7523
OKSAT-EN-EN-01-DN*	0.5374	0.5565	0.7344	0.6765	0.7523
UIOWA-EN-EN-03-DN	0.4990	0.5197	0.6254	0.6128	0.6998
UIOWA-EN-EN-02-DN	0.4955	0.5134	0.6153	0.5985	0.6919
BRKLY-JA-EN-01-DN	0.4874	0.5035	0.6072	0.5950	0.6891
UIOWA-EN-EN-04-DN	0.4869	0.5069	0.6110	0.5939	0.6889
SINAIUJAEN-EN-EN-02-DN	0.4759	0.4983	0.6154	0.5831	0.6941
SINAIUJAEN-EN-EN-04-DN	0.4611	0.4898	0.5951	0.5746	0.6824
BRKLY-EN-EN-01-DN	0.4495	0.4713	0.5690	0.5538	0.6588
SINAIUJAEN-EN-EN-01-D	0.4341	0.4564	0.5403	0.5491	0.6587
OKSAT-EN-EN-04-DN	0.4325	0.4580	0.5781	0.5685	0.6724
SINAIUJAEN-EN-EN-03-D	0.4266	0.4514	0.5296	0.5438	0.6505
UIOWA-EN-EN-01-D	0.4164	0.4372	0.5275	0.5388	0.6425
BRKLY-EN-EN-01-D	0.4066	0.4246	0.4931	0.5013	0.6012
BRKLY-JA-EN-01-D	0.3967	0.4081	0.4737	0.4739	0.5593
SJTUBCMI-EN-EN-02-DN	0.3648	0.3884	0.5127	0.4977	0.6045
OKSAT-EN-EN-05-DN	0.3334	0.3512	0.5142	0.4690	0.5862
SJTUBCMI-EN-EN-01-DN	0.3326	0.3557	0.4498	0.4511	0.5772
INESCID-EN-EN-01-D	0.3260	0.3497	0.4591	0.4563	0.5791
INESCID-EN-EN-03-D	0.3200	0.3362	0.4499	0.4550	0.5224
SJTUBCMI-EN-EN-04-D	0.3141	0.3370	0.4429	0.4523	0.5567
IRNLP-EN-EN-03-DN	0.3128	0.3354	0.4343	0.4281	0.5544
IRNLP-EN-EN-04-DN	0.3123	0.3351	0.4358	0.4270	0.5538
INESCID-EN-EN-04-D	0.3027	0.3183	0.4457	0.4378	0.5110
IRNLP-EN-EN-01-D	0.2999	0.3242	0.4257	0.4237	0.5448
IRNLP-EN-EN-02-D	0.2981	0.3225	0.4257	0.4215	0.5430
SJTUBCMI-EN-EN-03-D	0.2885	0.3111	0.4217	0.4085	0.5338
RMIT-EN-EN-02-D	0.2721	0.2939	0.3427	0.3674	0.4916
RMIT-EN-EN-01-D	0.2382	0.2427	0.4181	0.3613	0.4165
GETUA-EN-EN-01-D	0.2026	0.2159	0.2903	0.3081	0.4063
INESCID-EN-EN-02-D	0.2006	0.2093	0.3805	0.3284	0.3684
GETUA-EN-EN-03-DN	0.1872	0.2021	0.2763	0.2786	0.3899
GETUA-EN-EN-02-DN	0.1610	0.1774	0.1712	0.2524	0.3629
GETUA-EN-EN-04-D	0.1547	0.1730	0.2204	0.2409	0.3623
GETUA-EN-EN-05-D	0.1536	0.1740	0.3238	0.2567	0.3858
INESCID-EN-EN-05-DN	0.1392	0.1474	0.3137	0.2493	0.3195

- manual run (human interaction in query formulation)

Table 1b: GeoTime English best team performance for description only runs

RUN	AP	RUN	Q	RUN	nDCG@1000
OKSAT-EN-EN-03-D*	0.5376†	OKSAT-EN-EN-03-D*	0.5562	OKSAT-EN-EN-03-D*	0.7523
SINAIUJAEN-EN-EN-01-D	0.4341	SINAIUJAEN-EN-EN-01-D	0.4564	SINAIUJAEN-EN-EN-01-D	0.6587
UIOWA-EN-EN-01-D	0.4164	UIOWA-EN-EN-01-D	0.4372	UIOWA-EN-EN-01-D	0.6425
BRKLY-EN-EN-01-D	0.4066	BRKLY-EN-EN-01-D	0.4246	BRKLY-EN-EN-01-D	0.6012
INESCID-EN-EN-01-D	0.3260	INESCID-EN-EN-01-D	0.3497	INESCID-EN-EN-01-D	0.5791

- manual run (human interaction in query formulation)

† statistically significant difference (α=0.05) from the value of the run in the next row

Table 2a: GeoTime Japanese mean performance for three performance metrics for 31 submitted runs

RunName	MAP	Q	nDCG@10	nDCG@100	nDCG@1000
OKSAT-JA-JA-02-DN*	0.6506	0.6708	0.7479	0.7922	0.8547
OKSAT-JA-JA-03-D*	0.6449	0.6666	0.7565	0.7895	0.8542
OKSAT-JA-JA-01-DN*	0.6426	0.6622	0.7455	0.7826	0.8503
OKSAT-JA-JA-04-DN*	0.5733	0.6080	0.7149	0.7400	0.8179
HU-KB-JA-JA-03-D	0.4490	0.4804	0.4383	0.5960	0.6630
HU-KB-JA-JA-01-D	0.4385	0.4666	0.4282	0.5672	0.6298
HU-KB-JA-JA-02-DN	0.4385	0.4666	0.4282	0.5672	0.6298
HU-KB-JA-JA-05-D	0.4368	0.4648	0.4284	0.5646	0.6273
OKSAT-JA-JA-05-DN	0.4281	0.4642	0.5878	0.6170	0.7237
KOLIS-JA-JA-05-D	0.4227	0.4540	0.4782	0.5602	0.6294
HU-KB-JA-JA-04-D	0.4108	0.4458	0.4173	0.5319	0.6085
KOLIS-JA-JA-03-D	0.3996	0.4279	0.4013	0.5222	0.6027
KOLIS-JA-JA-01-D	0.3860	0.4180	0.4847	0.5256	0.6111
KOLIS-JA-JA-02-D	0.3815	0.4178	0.4641	0.5206	0.6042
RMIT-JA-JA-01-D	0.3779	0.4119	0.4777	0.5196	0.6152
BRKLY-JA-JA-01-DN	0.3716	0.3836	0.4362	0.4844	0.5696
RMIT-JA-JA-04-D	0.3671	0.3714	0.5289	0.5304	0.5942
KOLIS-JA-JA-04-D	0.3502	0.3822	0.3713	0.4773	0.5715
RMIT-JA-JA-05-D	0.3376	0.3398	0.5035	0.4899	0.5556
RMIT-JA-JA-03-D	0.3282	0.3349	0.4920	0.4777	0.5529
RMIT-JA-JA-02-D	0.3084	0.3239	0.3580	0.4050	0.4695
BRKLY-EN-JA-01-DN	0.3081	0.3214	0.3733	0.4250	0.5151
NAK-JA-JA-01-D	0.2928	0.3190	0.3256	0.4017	0.4936
KOLIS-EN-JA-05-D	0.2837	0.3175	0.3260	0.3995	0.4971
KOLIS-EN-JA-02-D	0.2833	0.3212	0.3545	0.4064	0.5029
KOLIS-EN-JA-01-D	0.2799	0.3165	0.3383	0.4002	0.4950
KOLIS-EN-JA-03-D	0.2749	0.3065	0.2772	0.3805	0.4840
KOLIS-EN-JA-04-D	0.2679	0.3022	0.3001	0.3820	0.4842
BRKLY-JA-JA-01-D	0.2475	0.2640	0.3250	0.3492	0.4157

BRKLY-EN-JA-01-D	0.2398	0.2550	0.3124	0.3326	0.4211
NAK-JA-JA-02-D	0.2201	0.2489	0.2311	0.3271	0.4301

Table 2b: GeoTime Japanese best team performance for description only runs

RUN	AP	RUN	Q	RUN	nDCG@1000
OKSAT-JA-JA-03-D*‡	0.6449	OKSAT-JA-JA-03-D*	0.6666	OKSAT-JA-JA-03-D*	0.8547
HU-KB-JA-JA-03-D	0.4490	HU-KB-JA-JA-03-D	0.4804	HU-KB-JA-JA-03-D	0.6630
KOLIS-JA-JA-05-D	0.4227	KOLIS-JA-JA-05-D	0.4540	KOLIS-JA-JA-05-D	0.6294
RMIT-JA-JA-01-D	0.3779	RMIT-JA-JA-01-D	0.4119	RMIT-JA-JA-01-D	0.6152
NAK-JA-JA-01-D	0.2928	NAK-JA-JA-01-D	0.3190	NAK-JA-JA-01-D	0.4936

- manual run (human interaction in query formulation)

‡ statistically significant difference (α=0.01) from the value of the run in the next row

8. DISCUSSION

GeoTime for NTCIR-9 was the second evaluation of geotemporal information retrieval., i.e. queries or questions which have specific geographic and temporal content. Overall performance increased over the first GeoTime evaluation., and manual query construction by one group substantially enriched the relevance set. A variety of external resources such as Yahoo PlaceMaker, Wikipedia, DBpedia, Geonames Google Maps, and the Alexandria Digital Library gazetteer were used by the twelve participating groups. In this round, one topic was introduced which required reverse geocoding (i.e. finding place names from a latitude/longitude specification); this proved challenging for almost all groups.

One remaining task for the organizers is to undertake a detailed analysis of the relative importance of the geographic and temporal components in determining overall performance.

9. ACKNOWLEDGMENTS

We thank participant assessors for English from University of Alicante and University Jaen, SPAIN, Christopher Harris (University of Iowa), Yoonjae Jeong (Korea Advanced Institute for Science and Technology), Shane Culpepper and Falk Scholer (RMIT University, AUSTRALIA), in addition to the organizers.

10. REFERENCES

[1] S Asadi, C.-Y. Chang, X. Zhou, and J. Diederich. Searching the world wide web for local services and facilities: A review on the patterns of location-based queries. In W. Fan, Z. Wu, and J. Yang, editors, WAIM2005, pages 91–101. Springer LNCS 3739, 2005.

[2] F Gey, R Larson, N Kando, J Machado and T Sakai, NTCIR-8 GeoTime Overview, *Proceedings of NTCIR-8,* Tokyo, JAPAN, June, 2010.

[3] F Gey, R Larson, J Machado and M Yoshioka, A Microanalysis of Topic Variation for a Geotemporal Query, *Proceedings of EVIA-2011*, Tokyo JAPAN December 2011.

[4] L L Hill, *GeoReferencing: The Geographic Associations of Information*, MIT Press, Cambridge, MA 2006.

[5] C. B. Jones, A. I. Abdelmoty, D. Finch, G. Fu, and S. Vaid. The SPIRIT spatial search engine: architecture, ontologies and spatial indexing. In GiScience 2004, Oct. 2004, Adelphi, MD, pages 125–139, 2004. Cunningham,

[6] Larson, R Geographic information retrieval and spatial browsing. In GIS and Libraries: Patrons, Maps and Spatial Information, pages 81–124. UIUC - GSLIS, Urbana-Champaign, IL, 1996.

[7] R Larson, Probabilistic Text Retrieval for NTCIR9 GeoTime, In Proceedings of NTCIR-9 (this proceedings).

[8] I. Mani, J. Pustejovsky, and B. Sundheim. Introduction to the special issue on temporal information processing. ACM Transactions on Asian Language Information Processing (TALIP), 3(1):1–10, 2004.

[9] R. Purves, C. Jones, and P. Clough. GIR'10: 6th workshop on geographic information retrieval, 2010. http://www.geo.unizh.ch/ rsp/gir10/index.html.

[10] D Santos, L Cabral, GikiCLEF: Crosscultural issues in an international setting: asking non-English-centered questions to Wikipedia, CLEF 2009 Working Notes, http://www.clef-campaign.org/2009/working_notes/Santos-paperCLEF2009.pdf. September 2009, 21pp.

[11] Sakai, T et al Overview of NTCIR-8 ACLIA IR4QA, In the proceedings of NTCIR Workshop 8, Tokyo, JAPAN, June, 2010.

[12] Jannik Strötgen, Michael Gertz, and Pavel Popov. 2010. Extraction and exploration of spatio-temporal information in documents. In *Proceedings of the 6th Workshop on Geographic Information Retrieval* (GIR '10). ACM, New York, NY, USA, , Article 16 , 8 pages. DOI=10.1145/1722080.1722101 http://doi.acm.org/10.1145/1722080.1722101

Appendix: Four Most Difficult English Topics with Answer URLS


```
<TOPIC ID="GeoTime-0037">
  - <DESCRIPTION LANG="EN">
  - <![CDATA[
  What fatal accident occurred near (geographical
     coordinates 5°52′12″N
  5°45′00″E / 5.870°N 5.750°E / 5.870; 5.750), which
     killed hundreds of
  people, and when did it occur?
    ]]>
    </DESCRIPTION>
  - <DESCRIPTION LANG="JA">
  - <![CDATA[
  北緯 5 度 52 分 12 秒東経 5 度 45 分の近くで起きた数百人の死亡者
     を出した事故は、どのような事故ですか？また、それはいつ起き
     ましたか？
    ]]>
    </DESCRIPTION>
  - <NARRATIVE LANG="EN">
  - <![CDATA[
  This topic requires spatial reasoning, to look up
     places near the geographic
  coordinates and then search for the story about the
     accident which happened there.
    ]]>
    </NARRATIVE>
  - <NARRATIVE LANG="JA">
  - <![CDATA[
  このトピックでは、与えられた地理座標の近くの場所を探すという
     地理空間に関する推論を必要とし、その場所の近くで起きた事故
     について調べる。
    ]]>
    </NARRATIVE>
- <URLs>
  <URL
     LANG="EN">http://en.wikipedia.org/wiki/1998_Je
     sse_pipeline_explosion</URL>
    </URLs>
- <RELs>
  <REL LANG="JA">JA-981020058</REL>
  <REL LANG="JA">JA-981021059</REL>
    </RELs>
  <QUERYDATE YYYYMMDD="20051231" />
    </TOPIC>
  ===================================
<TOPIC ID="GeoTime-0044">
  - <DESCRIPTION LANG="EN">
  - <![CDATA[
  Describe when and where deadly earthquakes happened
     in South America?
    ]]>
    </DESCRIPTION>
  - <DESCRIPTION LANG="JA">
  - <![CDATA[
  南アメリカで起きた死者が出た地震について、いつ、どこで起きた
     かを述べよ。
    ]]>
    </DESCRIPTION>
  - <NARRATIVE LANG="EN">
  - <![CDATA[
  The user wants to know about earthquakes in which
     people died as a
  result. Where and when did such earthquakes occur
     in South America?]
    ]]>
    </NARRATIVE>
  - <NARRATIVE LANG="JA">
  - <![CDATA[
  ユーザは死者がでた地震について知りたい。南アメリカでこのよう
     な地震が、いつ、どこで、起きたかを知りたい。
    ]]>
    </NARRATIVE>
- <URLs>
  <URL
     LANG="EN">http://en.wikipedia.org/wiki/List_of_d
     eadly_earthquakes_since_1900</URL>
    </URLs>
- <RELs>
  <REL LANG="EN">682033</REL>
  <REL LANG="EN">681898</REL>
  <REL LANG="JA">JA-050927178</REL>
  <REL LANG="JA">JA-010627271</REL>
    </RELs>
  <QUERYDATE YYYYMMDD="20051231" />
    </TOPIC>
  ===================================
<TOPIC ID="GeoTime-0050">
  - <DESCRIPTION LANG="EN">
  - <![CDATA[
  When and where was CAFTA, the Central America Free
     Trade Agreement signed?
    ]]>
    </DESCRIPTION>
  - <DESCRIPTION LANG="JA">
  - <![CDATA[
  いつ、どこで、中米自由貿易協定（CAFTA）は署名されましたか？
    ]]>
    </DESCRIPTION>
  - <NARRATIVE LANG="EN">
  - <![CDATA[
  CAFTA, the Central America Free Trade agreement was
     signed by President Bush after controversial
     passage by the USA congress.  When and where did
     this signing take place?  The month and year are
     adequate answers as to when.
    ]]>
    </NARRATIVE>
  - <NARRATIVE LANG="JA">
  - <![CDATA[
  中米自由貿易協定（CAFTA）は、アメリカ議会における論争の末、ブ
     ッシュ大統領によって署名されました。いつ、どこで、この署名
     は行われましたか？正解には、月と年の情報があればよい。
    ]]>
    </NARRATIVE>
- <URLs>
```

```
  <URL
     LANG="EN">http://en.wikipedia.org/wiki/CAFTA</
     URL>
    </URLs>
- <RELs>
  <REL LANG="JA">JA-050804063</REL>
  <REL LANG="EN">NYT_ENG_20050802.0139</REL>
    </RELs>
  <QUERYDATE YYYYMMDD="20051231" />
    </TOPIC>

   ===================================
<TOPIC ID="GeoTime-0035">
  - <DESCRIPTION LANG="EN">
  - <![CDATA[
  When and where did a pipeline explosion occur in
     Africa killing over
  500 people?
    ]]>
    </DESCRIPTION>
  - <DESCRIPTION LANG="JA">
  - <![CDATA[
  500 人以上の死者を出したパイプライン事故は、アフリカのどこで
     、いつ起きましたか？
    ]]>
    </DESCRIPTION>
  - <NARRATIVE LANG="EN">
  - <![CDATA[
  An oil pipeline exploded in an African oil-
     producing country and the resulting fire killed
     more than 500 people.  The user wants to know
     where this took place and when was the date of
     the accident.
    ]]>
    </NARRATIVE>
  - <NARRATIVE LANG="JA">
  - <![CDATA[
  アフリカの産油国で起きたパイプラインの爆発で 500 人以上の死者
     を出す火災が起きた。ユーザはこの爆発が起きた場所と日付を知
     りたい。
    ]]>
    </NARRATIVE>
- <URLs>
                                                 <URL
     LANG="EN">http://en.wikipedia.org/wiki/1998_Je
     sse_pipeline_explosion</URL>
    </URLs>
- <RELs>
  <REL LANG="JA">JA-981020058</REL>
  <REL LANG="JA">JA-981021059</REL>
    </RELs>
  <QUERYDATE YYYYMMDD="20051231" />
   </TOPIC>
```