Abstract. This paper presents a collaborative semantic tracking framework based on topic maps which aims to integrate and organize the data/information resources that spread throughout the Internet in the manner that makes them useful for tracking events such as natural disaster, and disease dispersion. We present the architecture we defined in order to support highly relevant semantic management and to provide adaptive services such as statistical information extraction technique for document summarization. In addition, this paper also carries out a case study on disease dispersion domain using the proposed framework.

1 Introduction

This paper gives an overview of a generic architecture we are currently building as part of the Semantic Tracking project in cooperation with the FAO AOS project [32]. Initiated by FAO and KU, the Semantic Tracking project aims at providing a wide area collaborative semantic tracking portal for monitoring important events related to agriculture and environment, such as disease dispersion, flooding, or dryness. This implies to deal with any kind of multilingual internet news and other online articles (e.g. wiki-like knowledge and web logs); it describes the world around us rapidly by talking about the update events, states of affairs, knowledge, people and experts who participate in.

Therefore, the Semantic Tracking project targets to provide adaptive services to large group of users (e.g. operator, decision makers), depending on all the knowledge we have about the environment (users themselves, communities they are involved in, and device he’s using). This vision requires defining an advanced model for the classification, the evaluation, and the distribution of multilingual semantic resources. Our approach fully relies on state of the art knowledge management strategies. We define a global collaborative architecture that allows us to handle resources from the gathering to the dissemination. However, sources of these data are scattered across several locations and web sites with heterogeneous formats that offer a large volume of unstructured information. Moreover, the needed knowledge was too difficult to find since the traditional search engines return ranked retrieval lists that offer little or no information on semantic
relationships among those scattered information, and, even if it was found, the located information often overload since there was no content digestion. Accordingly, the automatic extraction of information expressions, especially the spatial and temporal information of the events, in natural language text with question answering system has become more obvious as a system that strives for moving beyond information retrieval and simple database query.

However, one major problem that needs to be solved is the recognition of events which attempts to capture the richness of event-related information with their temporal and spatial information from unstructured text. Various advanced technologies including name entities recognition and related information extraction, which need natural language processing techniques, and other information technologies, such as geoinformation processing, are utilized part of emerging methodologies for information extraction and aggregation with problem-solving solutions (e.g. “know-how” from livestock experts from countries with experiences in handling bird flu situation). Furthermore, ontological topic maps are used for organizing related knowledge.

In this paper, we present our proposal aiming to integrate and organize the data/information resources dispersed across web resources in a manner that makes them useful for tracking events such as natural disaster, and disease dispersion. The remainder of this paper is structured as follows: Section 2 describes the key issues in information tracking as nontrivial problems; In Section 3 we introduce the framework architecture and its related many-sorted algebra. Section 4 gives more details of the system process regarding the information extraction module. Section 5 discusses the personalized services (e.g. knowledge retrieval service and visualization service) provided for collaborative environments. Finally, in Section 6, we conclude and give some forthcoming issues.

2 Key Issues of Information Tracking

Collecting and extracting data from the Internet have two main nontrivial problems: overload and scattered information, and salient information and semantic extraction from unstructured text. Many experiences [20, 21, and 35] have been done regarding event tracking or special areas or areas related to events monitoring (e.g. the best practice for governments to handle bird flu situation), the collection of important events and their related information (e.g. virus transmission from one area to other locations and from livestock to humans).

Firstly, target data used for semantic extraction are organized and processed to convey understanding, experience, accumulated learning, and expertise. However, sources of these data are scattered across several locations and websites with heterogeneous formats. For example, the information about Bird Flu consisting of policy for controlling the events, disease infection management, and outbreak situation may appear in different websites as shown in Fig. 1.
Consequently, collecting required information from scattered resources is very difficult since the semantic relations among those resources are not directly stated. Although it is possible to gather those information, the collected information often overload since there is no content digestion. Accordingly, solving those problems manually is impossible. It will consume a lot of time and CPU power. The system that can collect, extract and organize those information according to contextual dimensions automatically, is our research goal for knowledge construction and organization.

Secondly, only salient information must be extracted to reduce time consumption for users to consume the information. In many case, most of salient information (e.g. time of the event, location that event occurred, the detail of the event) are left implicitly in the texts.

For example: in the text in Fig. 1, the time expression “15 February” mentioned only “date and month” of the bird flu event but did not mention the ‘year’. The patient and her condition (i.e. ‘37-year-old female’, and ‘died’) was caused by bird flu which is written in the text as ‘Avian influenza’ and ‘H5N1 avian influenza’. Accordingly, the essential component of computational model for event information capturing is the recognition of interested entities including time expression, such as ‘yesterday’, ‘last Monday’, and ‘two days before’, which becomes an important part in the development of more robust intelligent information system for event tracking.

Information extraction in traditional way processes a set of related entities in the format of slot and filler, but the description of information in Thai text such as locations, patient’s condition, and time expressions can not be limited to a set of related entities because of the problems of using zero anaphora [17]. Moreover, to activate the frame for filling the information, name entity classification must be robust as it has been shown in [5].
3 A P2P Framework of Information Extraction for Event Tracking

In this section, we give an overview of the modeling we are providing. Preliminary parts of our framework have been previously introduced to the Natural Language Processing and Database community [15]. In the following, we present our P2P framework and related many-sorted algebra modeling.

3.1 Our Framework’s Semantic Tracking Algebra

Let us introduce our design approach of an ontological topic map for event semantic tracking. The ontological topic map [22] helps to establish a standardized, formally and coherently defined classification regarding event tracking.

One of our current focus and challenges has to develop a comprehensive ontology, which defines the terminology set, data structure and operations regarding semantic tracking and monitoring in the field of agriculture and environment.

The Semantic Tracking Algebra is a formal and executable instantiation of the resulting event tracking ontology. Our algebra has to achieve two tasks: (1) first, it serves as a knowledge layer between the users (e.g. agriculture experts) and the system administration (e.g. IT scientists and researchers).

Let us remind the notion of many sorted algebra [13]. Such algebra consists of several sets of values and a set of operations (functions) between these sets. Our Semantic Tracking Algebra is a domain-specific many-sorted algebra incorporating a type system for agriculture and environment data. It consists of two sets of symbols called sorts (e.g. topic, RSS postings) and operators (e.g. tm_transcribe, semantic_similarity); the function sections constitute the signature of the algebra. Its sorts, operators, sets, and functions are derived from our agriculture ontology. Second order signature [14] is based on two coupled many-sorted signatures where the top-level signature provides kinds (set of types) as sorts (e.g. DATA, RESOURCE, SEMANTIC_DATA) and type constructors as operators (e.g. set).

To illustrate the approach, we assume the following simplified many-sorted algebra:

**Kinds** DATA, RESOURCE, SEMANTIC_DATA, TOPIC_MAPS, SET

**Type**

| Constructor | Value |
|-------------|-------|
| DATA        | topic |
| RESOURCE    | rss, htm // resource document type |
| SEMANTIC_DATA | lsi_sm, rss_sm, htm_sm // Semantic and metadata vectors |
| TM          | tm(topic maps) |

**Unary operations**

\[ \forall \text{resource in RESOURCE, } \text{resource } \rightarrow \text{sm: SEMANTIC_DATA, tm } \text{tm_transcribe} \]

\[ \forall \text{sm in SEMANTIC_DATA, } \text{sm } \rightarrow \text{set(tm)} \]

\[ \text{semantic_similarity} \]
The notion sm:SEMANTIC_DATA is to be read as “some type sm in SEMANTIC_DATA,” and means there is a typing mapping associated with the tm_transcribe operator. Each operator determines the result type within the kind of SEMANTIC_DATA, depending on the given operand resource types.

**Binary operations**

\[ \forall \text{tm in TOPIC_MAPS}, \ (\text{tm})' \rightarrow \text{tm} \quad \text{topicmaps\_merging} \]

\[ \forall \text{sm in SEMANTIC\_DATA}, \ \forall \text{tm in TOPIC\_MAPS}, \quad \text{sm,tm} \rightarrow \text{tm} \quad \text{semantic\_merging} \]

\[ \forall \text{topic in DATA}, \ \forall \text{tm in TOPIC\_MAPS}, \ 
\quad \text{set(tm)} \times (\text{topic} \rightarrow \text{bool}) \rightarrow \text{set(tm)} \quad \text{select} \]

The semantic merging operation takes two or more operands that are all topic maps values. The select takes an operand type set (tm) and a predicate of type topic and returns a subset of the operand set fulfilling the predicate. From the implementation of view, the resource algebra is an extensible library package providing a collection of resource data types and operations for agriculture and environment resource computation. The major research challenge will be the formalization and the standardization of cultural resource data types and semantic operations through ISO standardization.

**Fig 2. Architecture of our proposed P2P Semantic Tracking System**
3.2 Framework of P2P Semantics Tracking System

As shown in Fig. 2, the proposed framework consists of six main services. The detail of each service is outlined as followed:

**Information and Knowledge Extraction and Management Services:** To generate useful knowledge from collected documents, two important modules, information extraction and knowledge extraction, are utilized. Ontological topic maps and domain-related ontologies defined in OWL [9] are used as a knowledge base to facilitate the knowledge construction and storage process as it has been shown in Garsho’s review [11]. The TOPIC MAPS ISO standard (ISO13250) facilitates the knowledge interoperability and composition. The information extraction and integration module is responsible for summarizing the document into a predefined frame-like/structured database, such as <disease name, dispersion location and time, status of patient’s condition>. The knowledge extraction and generalization is responsible for extracting useful knowledge (e.g. general symptom of disease) from collected document. Latent semantic analysis will be applied to find new knowledge or relationships that are not explicitly stored in the knowledge repository. Language engineering and knowledge engineering techniques are key methods to build the target platform. For language engineering, word segmentation [31], named entity recognition [6], shallow parsing [28], shallow anaphora resolution and discourse processing [6, 7, and 12] have been used. For knowledge engineering, ontological engineering, task-oriented ontology, ontology maintenance [16] and Topic Maps [5] model have been applied.

**Distributed Information Collection Management Service:** The information, both unstructured and semi-structured documents are gathered from many sources. Periodic web crawler and HTML Parser [33] are used to collect and organize related information. The domain specific parser [17] is used to extract and generate meta-data (e.g. title, author, and date) for interoperability between disparate and distributed information. The output of this stage is stored in the document warehouse.

**Content-based Metadata Extraction Service:** To organize the information scattered at several locations and websites, Textual Semantics Extraction [27] is used to create a semantic metadata for each document stored in the document warehouse. Guided by domain-based ontologies associated to reasoning processes [23] and Ontological topic map, the extraction process can be taught of as a process for assigning a topic to considered documents or extracting contextual metadata from documents following Xiao’s approach [36].

**Knowledge Retrieval Service:** This module is responsible for creating response to users’ query. The query processing based on TMQL-like requests is used to interact with the Knowledge management layer.

**Knowledge Visualization:** After obtaining all required information from the previous module, the last step is to provide the means to help users consume that information in an efficient way. To do this, many visualization functions is provided. For example,
Spatial Visualization can be used to visualize the information extracted from the Information Extraction module and Graph-based Visualization can be used to display hierarchical categorization in the topic maps in an interactive way [27]. Due to page limitation, this paper will focus on only Information Extraction module, Knowledge Retrieval Service module and Knowledge Visualization Service module.

4 Information Extractions

The proposed model for extracting information from unstructured documents consists of three main components, namely Entity Recognition, Relation Extraction, and Output Generation, as illustrate in Fig. 3. The Entity Recognition module is responsible for locating and classifying atomic elements in the text into predefined categories such as the names of diseases, locations, and expressions of times. The Relation Extraction module is responsible for recognizing the relations between entities recognized by the Entity Recognition module. The output of this step is a graph representing relations among entities where a node in the graph represents an entity and the link between nodes represents the relationship of two entities. The Output Generation module is responsible for generating the n-tuple representing extracted information from the relation graph. The details of each module are described as followed.

4.1 Entity Recognition

To recognize an entity in the text, the proposed system utilizes the work of H. Chanlekha and A. Kawtrakul [6] that extracts entity using maximum entropy [2], heuristic information and dictionary. The extraction process consists of three steps. Firstly, the candidates of entity boundary are generated by using heuristic rules, dictionary, and statistic of word co-occurrence. Secondly, each generated candidate is then tested against the probability distribution modeled by using maximum entropy. The features used to model the probability distribution can be classified into four categories: Word Features, Lexical Features, Dictionary Features, and Blank Features as described in [7]. Finally, the undiscovered entity is extracted by matching the extracted entity against the rest of the document. The experiment with 135,000 words corpus, 110,000 words for training and 25,000 words for testing, shown that the precision, recall and f-score of the proposed method are 87.60%, 87.80%, 87.70% respectively.

4.2 Relation Extraction

To extract the relation amongst the extracted entities, the proposed system formulates the relation extraction problem as a classification problem. Each pair of extracted entity is tested against the probability distribution modeled by using maximum entropy to determine whether they are related or not. If they are related, the system will create an edge between the nodes representing those entities. The features used to
model the probability distribution are solely based on the surface form of the word surrounding the considered entities; specifically, we use the word n-gram and the location relative to considered entities as features. The surrounding context is classified into three disjointed zone: prefix, infix, and suffix. The infix is further segmented into smaller chunks by limiting the number of words in each chunk. For example, to recognize the relation between VICTIM and CONDITION in the sentence “The [VICTIM] whose [CONDITION] was announced on ....”, the prefix, infix and suffix in this context is ‘the’, ‘whose’, and ‘was announced on ....’ respectively.

Fig. 3. Overview of the information extraction module

To determine and to assess the “best” n-gram parameter and number of words in each chunk of the system, we conduct the experiment with 257 documents, 232 documents for training and 25 documents for testing. We vary the n-gram parameter from 1 to 7 and set the number of words in each chunk as 3, 5, 7, 10, 13, and 15. The result is illustrated in Fig. 4. The evident shows that f-score is maximum when n-gram is 4 and number of words in each chunk is 7. The precision, recall and f-score at the maximum f-score are 58.59%, 32.68% and 41.96% respectively.

4.3 Output Generation

After obtaining a graph representing relations between extracted entities, the final step of information extraction is to transform the relation graph into the n-tuple representing extracted information. Heuristic information is employed to guide the transformation process. For example, to extract the information about disease outbreak (i.e. disease name, time, location, condition, and victim), the transformation process will starts by analyzing the entity of the type condition, since each n-tuple can
contain only one piece of information about the condition. It then travels the graph to obtain all entities that are related to considered condition entity. After obtaining all related entities, the output n-tuple is generated by filtering all related entities using constrain imposed by the property of each slot. If the slot can contains only one entity, the entity that has the maximum probability will be chosen to fill the slot. In general, if the slot can contain up to n entities, the top-n entities will be selected. In addition, if there is no entity to fill the required slot, the mode (most frequent) of the entity of that slot will be used to fill instead. The time expression normalization using rule-based system and synonym resolution using ontology are also performed in this step to generalize the output n-tuple. The example of the input and output of the system are illustrated in Fig. 3.

Fig. 4. Experimental results of relation extraction module (the legend in the figure is the number of words in each chunk)

5 Adaptive Services

Distributed adaptive and automated services require exploiting all the environmental knowledge stored in ontological topic maps that is available about the elements involved in the processes [24]. An important category of this knowledge is related to devices’ states; indeed, knowing if a device is on, in sleep mode, off, if its battery still has autonomy of five minutes or four days, or if it has a wired or wireless connection, etc. helps adapting services that can be delivered to this device. For each device, we consider a state control that is part of the device’s profile. Then, of course, we use the information contained in communities’ and users’ profiles. Personalized services rely on user-related contexts such as localization, birth date, languages abilities, professional activities, hobbies, communities’ involvement, etc. that give clues to the system about users’ expectations and abilities. In the remainder of this section, we present the two main adaptive services: the knowledge service and the knowledge visualization service based on our model.
5.1 Knowledge Retrieval Service

The Knowledge Retrieval Service module is responsible for interacting with the topic maps repository to generate answers to user’s TMQL-like queries [33]. The framework currently supports three types of query. The detail of each query type is summarized in Table 1.

| Query type          | Description                                                                 |
|---------------------|-----------------------------------------------------------------------------|
| Query by Object     | A mechanism employed when users know the object but want to acquire more information/knowledge about it. The query example is as following:  
  ```sql
  SELECT qa_who, lblWho  
  FROM {qa_who} ne:text {lblWho}  
  WHERE (lblWho like "*/afii59744;/afii59700;/afii59751;/afii59681")  
  USING NAMESPACE ne = <http://naist.cpe.ku.ac.th/EventTracking#>
  ``` |
| Query by Relationship | A mechanism employed when users know the relation label. For example, user can access knowledge repository such as "ne:atLocation". The query example is as following:  
  ```sql
  SELECT disease, lblDisease, location, lblLocation  
  FROM {disease} ne:text {lblDisease}, {disease} ne:atLocation {location}, {location} ne:text {lblLocation}  
  WHERE (lblDisease like "*/afii59723;/afii59719;/afii59728;/afii59700;/afii59705;/afii59681")  
  AND (lblLocation like "*/afii59744;/afii59719;/afii59733;/afii59714;/afii59700;")  
  USING NAMESPACE ne = <http://naist.cpe.ku.ac.th/EventTracking#>
  ``` |
| Query by Instance   | A mechanism employed when users know the instance or some parts of instance label that can access to knowledge repository such as "ne:Disease-3-10". The query example is as following:  
  ```sql
  SELECT disease, lblDisease, location, lblLocation  
  FROM {disease} ne:text {lblDisease}, {disease} ne:atLocation {location}, {location} ne:text {lblLocation}  
  WHERE (lblDisease like "*/afii59723;/afii59719;/afii59728;/afii59700;/afii59705;/afii59681")  
  AND (lblLocation like "*/afii59744;/afii59719;/afii59733;/afii59714;/afii59700;")  
  USING NAMESPACE ne = <http://naist.cpe.ku.ac.th/EventTracking#>
  ``` |

5.2 Knowledge Visualization Service

The Knowledge Visualization Service is responsible for representing the extracted information and knowledge in an efficient way. Users require to access to concise organization of the knowledge. Schneiderman in [12] pointed that "the visual information-seeking mantra is overview first, zoom and filter, then details on-demand". In order to locate relevant information quickly and explore the semantic-related structure, our flexible approach regarding two ways of visualizations (spatial-based or graph-based visualization) is described in the following.
Spatial Visualization

The spatial-based visualization functions help users to visualize the extracted information (e.g. the bird flu outbreak situation extracted in Fig. 3.) using web-based geographical information system, such as Google Earth. This kind of visualization allows the users to click on the map to get the outbreak situation of the area according to their requests. In addition, by viewing the information in the map users can see the spatial relations amongst the outbreak situations easier than without the map. One usage example of Google Earth integrated system for visualizing the extracted information about bird flu situation is shown in Fig. 5.

![Google Earth visualization for bird flu outbreak tracking](image)

Graph-based Visualization

The graph-based visualization function is useful to show the global structure of topic maps and relations between different nodes in a 3D visual space. End-users can directly select the requested topics and related associations. The graph viewer provides a better global understanding of the content by exploring through graph nodes. The kind of intuitive visualization of topic maps allows browsing through all the topics and related relationships defined in the topic map as shown in Fig. 6. The graph can be moved and restructured along its topological view according to the user’s need and selections.
6 Related Works

6.1 Distributed Services

We agree that distributed knowledge management has to assume two principles [4] related to the classification: (1) autonomy of classification for each knowledge management unit (such as community), and (2) coordination of these units in order to ensure a global consistency. Having a decentralized peer-to-peer knowledge management, the SWAP platform [25] is designed to enable knowledge sharing in a distributed environment. Pinto et al. provide interesting updates and changes support between peers. However, vocabularies in SWAP have to be harmonized; which implies to have some loss of knowledge consistency. But even if we share the approach of core knowledge structure that is expendable, the vocabulary, in our case, is common and fully shared by the community, so the knowledge evaluation and comparison can be more effective. Moreover, SWAP provides some kind of personalization (user interface mainly) but does not go as far as the Semantic Tracking does. From our point of view, SWAP definitely lacks environmental knowledge management that is required to perform advanced services; on the other hand, DBGlobe [26] is a service-oriented peer-to-peer system where mobile peers
carrying data provide the base for services to be performed. Its knowledge structure is quite similar to our project as it is using metadata about devices, users and data within profiles; moreover, communities are also focused on one semantic concept. DBGlobe relies on AXML [3] in order to perform embedded calls to Web services within XML. Thus, it provides a very good support for performing services but does not focus on users and environments knowledge in order to offer optimized authoritarian adaptive services. Described as a P2P DBMS, AmbientDB [10] relies on the concept of Ambient Intelligence, which is very similar to our vision of adaptive services with automatic cooperation between devices and personalization. However, although AmbientDB is using the effective Chord Distributed Hash Table to index the metadata related to resources, it lacks the environmental knowledge management provided inside our project that is necessary to achieve adaptive collaborative distribution and personalized query optimization.

6.2 Semantic Tracking

The extraction framework described in this paper is closely related to ProMED-PLUS [37], a system for the automatic “fact” extraction from plain-text reports about outbreaks of infectious epidemics around the world to database, and MiTAP [8], a prototype SARS detecting, monitoring and analyzing system. The difference between our framework and those systems is that we also emphasize on generating the semantic relations among the collected resources and organizing those information by using topic map model.

The proposed information extraction model that formulates the relation extraction problem as a classification problem is motivated by the work of J.Suzuki et. al. [30]. This innovated work has proposed a HDAG kernel solving many problems in natural language processing. The use of classification methods in information extraction is not new. Intuitively, one can view the information extraction problem as a problem of classifying a fragment of text into a predefined category which results in a simple information extraction system such as a system for extracting information from job advertisements [38] and business cards [19]. However, those techniques require the assumption that there should be only one set of information in each document, while our model could support more than one set of information.

8 Conclusion and Forthcoming Issues

As communities generate increasing amounts of transactions and deal with fast growing data, it is very important to provide new strategies for their collaborative management of knowledge. In this paper, we presented and described our proposal regarding Information Modeling for Adaptive Semantic Management which aims at extracting information and knowledge from unstructured documents that spread throughout the Internet by emphasizing on information extraction technique, event tracking and knowledge organizing. We first motivated the need for such modeling in order to provide personalized services to users who are involved in semantic tracking communities. The motivation for this work is definitely to improve user’s access to
semantic information and to reach high satisfaction levels for decision making. Then, we gave an overview of our approach’s algebra with its operators, focusing on update and consistency policies. We finally proposed and defined adaptive services that enable collaborative project to automatically dispatch semantic and to make the query results more relevant.

This challenging work needs more complicate natural language processing with deeply semantic relations interpretation.

Acknowledgement

The work described in this paper has been supported by the grant of National Electronics and Computer Technology Center (NECTEC) No. NT-B-22-14-12-46-06, under the project “A Development of Information and Knowledge Extraction from Unstructured Thai Document”.

References

1. Kongwan, A., Kawtrakul, A.: Know-what: A Development of Object Property Extraction from Thai Texts and Query System. In Proceedings of SNLP-2005. Bangkok, Thailand (2005) pp. 157-162. ISBN: 974-9942-52-3
2. Berger, A.L., Pietra, S.-A.D., Pietra, V.-J.D.: A maximum entropy approach to natural language processing. Computational Linguistics. Vol. 22 (1996) pp. 39-71
3. Biswas D., and Kim I. Atomicity for P2P based XML Repositories, in ICDE 2007.
4. Bonifacio M, Bouquet P., and Cuel R., “The Role of Classification(s) in Distributed Knowledge Management,” Proc. of KES, Podere d’Ombriano, Italy, September 16-18 2002.
5. Biezunski, M., Bryan, M., Newcomb, S.: ISO/IEC JTC1/SC34. 22 May 2002. Available at http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0322.htm
6. Chareonsuk, J., Kawtrakul, A.; Thai Named Entity Extraction by incorporating Maximum Entropy Model with Simple Heuristic Information. In Proceedings of IJCNLP-2004. Hainan, China (2004) pp. 49-55
7. Chareonsuk, J., Sukvakree, Y., Kawtrakul, A.: Elementary Discourse unit Segmentation for Thai using Discourse Cue and Syntactic Information. In Proceedings of SNLP. Bangkok, Thailand. (2005) pp. 85-90. ISBN: 974-9942-52-3
8. Damianos, L., Bayer, S., Chisholm, M.A., Henderson, J., Hirschman, L., Morgan, W., Ubaldino, M., Zarrella, J: MiTAP for SARS detection. In Proceedings of the Conference on Human Language Technology. Boston, USA (2004) pp. 241–244
9. Dean, M., Connolly, D., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L.A.: OWL Web Ontology Language Reference. W3C Recommendation, 10 Feb. 2004. Available at http://www.w3.org/TR/owl-ref/
10. Fontijn, W., Boncz, P.A., AmbientDB: P2P Data Management Middleware for Ambient Intelligence. In Proceedings of PerCom Workshops, USA, pp.203–207, March 14-17 2004.
11. Garshol, L.M.: Living with Topic Maps and RDF. May 2003. Available at http://www.idealliance.org/papers/dx_xml03/papers/02-03-06/02-03-06.html
12. Grosz, B., Joshi, A., Weinstein, S.: Centering: A Framework for Modeling the Local Coherence of Discourse. Computational Linguistics, Vol. 21 (1995) pp. 203-225
13. Güting, R. H. 1989. Gral: an extensible relational database system for geometric applications. In Proceedings of the 15th international Conference on Very Large Data Bases (Amsterdam, The Netherlands). Very Large Data Bases. Morgan Kaufmann Publishers, San Francisco, CA, 33-44.

14. Güting, R. H. 1993. Second-order signature: a tool for specifying data models, query processing, and optimization. In Proceedings of the 1993 ACM SIGMOD international Conference on Management of Data (Washington, D.C., United States, May 25 - 28, 1993). P. Buneman and S. Jajodia, Eds. SIGMOD ’93. ACM Press, New York, NY, 277-286. DOI= http://doi.acm.org/10.1145/170035.170079

15. Kawtrakul, A., Permpool, T., Yingsaeree, C., and Andres, F., A Framework of NLP based Information Tracking and related Knowledge Organizing with Topic Maps, in Z. Kedad et al. (Eds): NLDB2007, LNCS No. 4592 Natural Language Processing and Information Systems, Springer-Verlag, pp.272-283, ISBN 978-3-540-73350-8, 12th International Conference on Applications of Natural Language to Information Systems (NLDB 2007), June 27-29, 2007, CNAM, Paris, France

16. Kawtrakul, A., Suktarachan, M., Insombut, A.: Automatic Thai Ontology Construction and Maintenance System. In Proceedings of Ontolex Workshop on LREC (2004) pp. 68-74

17. Kawtrakul, A., Yingsaeree, C.: A Unified Framework for Automatic Metadata Extraction from Electronic Document. In Proceedings of The International Advanced Digital Library Conference. Nagoya, Japan (2005)

18. Kongwan, A., Kawtrakul, A.: Know-what: A Development of Object Property Extraction from Thai Texts and Query System. In Proceedings of SNLP-2005. Bangkok, Thailand (2005) pp. 157-162, ISBN: 974-9942-52-3

19. Kushmerick, N., Johnston, E., McGuinness, S.: Information extraction by text classification. In Proceedings of IJCAI-2001 Workshop on Adaptive Text Extraction and Mining. (2001)

20. Li, B., Li, W., Lu, Q., and Wu, M. 2005. Profile-based event tracking. In Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Salvador, Brazil, August 15 - 19, 2005). SIGIR ’05. ACM Press, New York, NY, 631-632.

21. Michael T. Chan, Anthony Hoogs, Zhaohui Sun, John Schmiederer, Rahul Bhotika, Gianfranco Doretto, "Event Recognition with Fragmented Object Tracks," icpr, pp. 412-416, In Proceedings of 18th International Conference on Pattern Recognition (ICPR’06), 2006

22. Naito M., Andres, F.: Application Framework Based on Topic Maps. In Proceedings of TMRA-2005. Leipzig, Germany (2005) Lecture Notes in Computer Science, Volume 3873, Feb 2006, Pages 42 - 52, DOI 10.1007/11676904_4, Charting the Topic Maps Research and Applications Landscape: First International Workshop on Topic Map Research and Applications, TMRA 2005, Leipzig, Germany, October 6-7, 2005, Revised Selected Papers Editors: Lutz Maicher, Jack Park ISBN: 3-540-32527-1

23. Pan, J.Z.: A Flexible Ontology Reasoning Architecture for the Semantic Web. IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 2, pp. 246-260, Feb., 2007

24. Perich, F., Joshi, A., Finin, T.W., Yesha, Y., “On Data Management in Pervasive Computing Environments,” Trans. Knowl. Data Eng., vol.16, no.5, pp.621–634, 2004.

25. Pinto H.S, Staab S., Sure S., and Tempich C., “OntoEdit Empowering SWAP: a Case Study in Supporting Distributed, Loosely-Controlled and evolving Engineering of oNTologies (DILIGENCE),” Proc. of EWS, Heraklion, Greece, pp.16–30, May 10-12 2004.

26. Pitoura, E., Abiteboul, S., Pfefer, D., Samaras, G., and Vazirgiannis, M. 2003. DBGlobe: a service-oriented P2P system for global computing. SIGMOD Rec. 32, 3 (Sep. 2003), 77-82.
27. Rajbhandari, S., Andres, F., Naito, M., Wuwongse, V.: Topic Management in Spatial-Temporal Multimedia Blog. In the 1st IEEE International Conference on Digital Information Management (ICDIM 2006), Bangalore, India, December, 6-8, 2006, pp. 81-88, ISBN:1-4244-0682-X

28. Satayamas, V., Thumkanon, C., Kawtrakul, A.: Bootstrap Cleaning and Quality Control for Thai Tree Bank Construction. In Proceedings of NCSEC-2005. Bangkok Thailand (2005) pp. 849-860, ISBN: 974-677-541-3

29. Shneiderman, B., 1996, The eyes have it: a task by data type taxonomy for information visualizations, Proceedings of 1996 IEEE Visual Languages, Boulder, CO, pp. 336-343.

30. Suzuki, J., Sasaki, Y., Maeda, E.: Kernels for structured natural language data. In Proceeding of NIPS 2003. (2003)

31. Sudprasert, S., Kawtrakul, A.: Thai Word Segmentation based on Global and Local Unsupervised Learning. In Proceedings of NCSEC-2003. Chonburi, Thailand (2003) ISBN: 974-3826-04-1

32. The AOS project, FAO. http://www.fao.org/aims/aos.jsp

33. Thamvijit, D., Chanlekha, H., Sirigayon, C., Permpool, T., Kawtrakul, A.: Know-who: Person Information from Web Mining. In Proceedings of NCSEC. Bangkok, Thailand (2005) pp. 849-860, ISBN: 974-677-541-3

34. Topic Map Query Language (TMQL) http://www.isotopicmaps.org/tmql/

35. Vargas-Vera, M., Celjuska, D. "Event Recognition on News Stories and Semi-Automatic Population of an Ontology," wi, pp. 615-618, IEEE/WIC/ACM International Conference on Web Intelligence (W104), 2004

36. Xiao, R., Tang, S., Li, L., Fang, L., Xu, Y., Chen, W., Xu, Y.: Using Categorial Context-SHOIQ(D) DL to Integrate Context-aware Web Ontology MetaData. In Proceedings of Second International Conference on Semantics, Knowledge, and Grid (SKG’06), 2006

37. Yangarber, R., Jokipii, L., Rauramo, A., Huttunen, S.: Information Extraction from Epidemiological Reports. In Proceedings of HLT/EMNLP-2005. Canada (2005)

38. Zavrel, J., Berck, P., Lavrijssen, W.: Information extraction by text classification: Corpus mining for features. In Proceedings of the workshop Information Extraction meets Corpus Linguistics. Athens, Greece (2000)