Classification of Web Resident Sensor Resources using Latent Semantic Indexing and Ontologies

Wabo Majavu and Terence van Zyl
Meraka Institute, CSIR
P.O. Box 395
Pretoria 0001, South Africa
(wmajavu/tvzyl@csir.co.za)

Tshilidzi Marwala
School of Electrical and Information Engineering
University of the Witwatersrand
Private Bag 3, Wits 2050
Johannesburg, South Africa
(t.marwala@ee.wits. ac.za)

***Abstract*—Web resident sensor resource discovery plays a crucial role in the realisation of the Sensor Web. The vision of the Sensor Web is to create a web of sensors that can be manipulated and discovered in real time. A current research challenge in the sensor web is the discovery of relevant web sensor resources. The proposed approach towards solving the discovery problem is to implement a modified Latent Semantic Indexing by making use of an ontology for classifying Web Resident Resources found in geospatial web portals. The paper presents the use of Latent Semantic Indexing, an information retrieval mechanism, biased by combining Ontology concepts to the terms and objects, for improving the knowledge extraction from web resident documents. The use of an Ontology, before indexing of terms, to create a semantic link between documents with relevant content improves automatic content extraction and document classification.**

***Index Terms*—Resource Classification, Document clustering, Latent Semantic Indexing, Ontolgies, Sensor Web.**

I. Introduction

Easy and fast access to domain specific information available on the World Wide Web (WWW) requires structuring of web resources[1]. Large databases containing digital geospatial data make use of the Internet as a platform for distributing these data sets. The growing volumes of sensor data sets published on the Internet are difficult for users to locate and access because of their unstructured nature[2]. Intelligent systems are required to address the challenge of relevant data discovery in large data sets made available in such a dynamic environment as the internet. Information Retrieval techniques have been explored in data mining research to enhance user searches and structuring of documents based on content[3]. Latent Semantic Indexing (LSI) was proposed by Deerwester as a document search and retrieval algorithm that organises documents into a semantic structure by using higher order term-clustering of the terms contained in the document corpus[4]. The documents and indexed terms are represented in a semantic space and relationships are established between documents even if they have no common terms between them by stuydying the pattern of occurrence of terms. It is proven that by using LSI for document content retrieval and clustering, some of the noise is removed,which is caused by terms occuring frequently next to different neighbouring terms, by the latent space representation. The structuring of documents in this manner allows for synonyms to be grouped together because of the underlying pattern of usage of terms. LSI has been implemented successfully in various domains as a semantic annotation, indexing, document retrieval tool[5]. The paper presents work which extends the semantic richness of LSI by combining its information retrieval ability with an ontology. The introduction of an ontology into LSIS introduces a bias towards the retrieval of web resources that have geospatially relevant criteria. The paper describes the use of the latent semantic vector space modelling of documents and semantically enhancing the clustering by using the existing relationships between concepts, specified in an ontology. The strength of LSI lies in redistribution of weights across connectivity paths established between documents[6]. The use of an ontology reenforces connections established between terms occurring in separate documents by virtue of a known semantic connection. The ontology serves as a representation of prior clustering knowledge. Using a term-relationship-term tripple represented in a knowledge base allows the automation of geospatial web document clustering as it formalises the logic representation of concepts and the semantic links between them. Document clustering is implemented using the proposed method in geospatial web portals in order to group the documents according to the feature observed by the web resource.

The remainder of this paper is organised as follows: In section II the Sensor Web is discussed. Section III gives an overview of web resource classification and some of its uses, leading to a description of LSI. Section IV covers how ontologies have been used for semantic information retrieval and section V discusses the experimental steps of the work done. A presentation of the results follows in sections VI and finally the conclusion and suggestions for future work, is presented in section VII.

II. Sensor Web

The Sensor Web is defined as a web of interconnected sensors which are fronted by interoperable Web Services that comply with standard specifications[7]. The Open Geospatial Consortium (OGC) has proposed a framework of open standards to help enable access and publishing of sensor data across different platforms worldwide as one of the efforts to enabling the creation of the Sensor Web. The realisation of the

1-4244-2384-2/08/$20.00

Sensor Web will facilitate internet accessibility and task-ability of heterogeneous web resident sensor resources (WRSR). The Sensor Web operates as an autonomous macro-instrument for environmental data gathering and processing which allows for re-use of this collected data for different purposes[8]. One of the functions of the Sensor Web is to allow the discovery of sensor systems and sensor observations filtered according to requirements that are entered by a user[9][10]. Activities that contribute towards the development of the Sensor Web aim to integrate and manage WRSR data. Discovery is targeted at these web resources.

III. Web Resource Classification

The rapid growth rate of internet resources make topic specific searches difficult. The rate at which resources are published is faster than the indexing mechanisms currently used[11]. The advantage of domain specific web portals is the single point of access they provide to web services, applications and data published by separate data providers[12]. Web portals are web sites that collect information on a common topic[13]. They provide links to other web resources and serves as a starting point to a selection of resources residing on the web. Even with integrated data access solutions such as portals, data clearing houses and catalogues, there is still a need for structured information representation to enable semantic searches of web resources[14]. Resource classification has been extensively researched in the field of Information Retrieval for search optimisation and logical organisation of web documents[15]. Techniques have been explored for WRSR classification, which looks at web page classification based on relevant content retrieval[16]. Automatic resource classification assists users in finding what they are looking for in large databases[17]. Classification of earth and space data by domain conceptualisation has been proposed to aid knowledge discovery in the large amounts of geospatial data available on the web[18]. Text categorisation is a method for classifying documents into predefined classes. A brief description of text categorisation using vector space representation is discussed and how semantic associations are introduced by LSI.

A. Vector Space Modeling

Vector space representation of documents involves modelling of the document corpus as a term-document matrix. Frequently occurring terms are indexed and arranged in the matrix, where for a matrix $\mathbf{X}$ with dimensions $(m \times n)$, rows represent documents and columns represent terms[19].

$$\mathbf{X} = \begin{bmatrix} d_1t_1 & d_1t_2 \cdots d_1t_n \\ d_2t_1 & d_2t_2 \cdots d_2t_n \\ \vdots & \vdots \\ d_mt_1 & d_mt_2 \cdots d_mt_n \end{bmatrix}$$

The matrix is a k-dimension space, where the values represent the weight of term n in document m. The arrangement of terms into vector space ignores grammatical word arrangement by forming a bag-of-words weighted by frequency of occurrence. Vector modelling is used, mostly in keyword searches where a user query is treated as a document with entered text as the terms and a match is made against the bag of words to return relevant documents.

B. Latent Semantic Indexing (LSI)

Latent Semantic Indexing reduces the vector space by creating a subspace of the matrix dimensions in order to remove noise and redundant terms. The reduced space presents a meaningful association between terms that in turn relate documents[3]. The first step is to index frequently occurring terms in a term-document matrix and compute singular value decomposition (SVD) from the original k-dimensional term-document matrix. SVD is a matrix decomposition method commonly used for data analysis. The original term-document matrix, $\mathbf{X}$, is decomposed into several matrices so their features can be revealed, for example document-document relationships. The decomposition is expressed as,

$$\mathbf{X}(SVD) = \mathbf{T}_{t\times k} \cdot \mathbf{S}_{k\times k} \cdot \mathbf{D}_{k\times d}$$

where, T is a left singular vector representing a term by dimension matrix, S is a singular value dimension by dimension matrix and D is a right singular vector representing document by document matrix [6]. The decomposed matrices are then truncated to a dimension less than the original k-value and the orginal $\mathbf{X}$ matrix approximated in the reduced latent space which better represents semantic relationships between terms compared to the original k-dimension document space. A study of LSI indicates its dependency on term-frequency and on the higher level co-occurrence of terms in different documents. Studies also show that LSI ignores the class structure of concepts[17]. The paper presents a technique where LSI is combined with an ontology modeling geospatial concepts that captures class structure and forms links between concepts which may not be captured by LSI alone after the dimension reduction.

IV. Use of Ontologies in Information Retrieval

An ontology is a formal description of concepts and a relation between them, that represents an area of knowledge and is usually expressed in a logic-based language. Ontologies together with class instances form a knowledge base which models a domain[20]. Ontologies can be used to share a common understanding of a domain structure and can be used to improve Web-based applications by re-use of domain knowledge[13]]. Ontologies are used for conceptual modelling in order to define concepts and relationships between concepts with varying levels of expressiveness by allowing restrictions to be placed on property fillers[21]. Semantic technologies make use of Ontologies for knowledge sharing and reasoning[5]. Ontologies have received attention in the Semantic Web Community as a solution to overcome key-word based information retrieval, where domain-specific knowledge bases are created to facilitate semantic searches[22].

V. Using Ontologies and LSI to Bias Document Clustering

Dimension reduction in LSI, followed by document-document clustering ignores the semantic relationships between concepts. The infrequent occurrence of a term and any links to neighbouring terms may cause it to be discarded or receive a low weighting even though it is of significance in a particular domain[17]. A domain specific ontology is therefore combined with LSI in order to enhance knowledge extraction in a geospatial web portal. A small document set is created by representing documents using headings, for experimental purposes, as in the original LSI implementation by Deerwester. Four experiments are conducted:

LSI is first implemented on two separate document sets which are water based and Ice based. The document headings are altered slightly by introducing common term between them. The goal is to illustrate the term co-occurrence dependency in order for links to be established between the two sets of documents. The Ontology headings are introduced, representing a third document set which is used to introduce additional links between concepts and bias LSI to the classification of WRSR. The fourth set of headings represent documents with content relating to jewellery and the word 'ice' is used in a different context. This is to illustrate that by using an ontology to establish additional links between the other three document sets, documents with irrelevant information but having the same key-words occurring in them receive a weaker relationship with the rest of the corpus. The ontology water-based headings used are extracted from a constructed WRSR ontology described below and the results are presented.

1) WRSR Ontology Description: The ontology aims to profile a sensor resource, using possible classes and associations, and express common elements between sensor resources to enable unsupervised identification on the Web. The web resources of interest are those which contain geospatial data or links to sensor generated geospatial data. Resource Content is classified according to the property being observed by the sensor. Sensor can refer to an individual sensor or a sensor network that produces a stream of time-stamped data or measurements of observed features. The WRSR ontology contains a description of the hydrosphere concept adopted from the upper ontology for Earth System Science, provided by Semantic Web for Earth and Environmental Terminology (SWEET). The SWEET subclasses structure knowledge about the different forms that water can represent itself, either in the atmosphere or on the earth surface. The test ontology in this paper was constructed for experimental reasons as a demonstration of the deductive reasoning over a sample representation of the hydrosphere concept and not for complete hydrosphere concept modelling.

VI. Results

The results presented are first of the term-term matrix in order to show the weighting of relationships between two indexed terms. Term-document relationship is then plotted for each of the experiments for the purpose of document clustering visualisation. A small document set is used for classification, therefore we adhere to k=2 dimension reduction to approximate the original document-term matrix. k Is the maximum vector space represented by the term document matrix and can be reduced to a value less than the original maximum value in order to create a latent space. Empirically determining the k-value was considered but it is suggested in related literature that this should be done empirically on a large corpus where values of k can be approximated to 1000[6]. In our implementation largest possible value for k is 8, which is the matrix rank of the singular diagonal matrix after matrix decomposition. Fig. 1 shows the partition between terms t1t4 (water related) and t5-t9 (ice related). Water related headings have no x-component and Ice related headings have no y-component. The only relationship determined by LSI is term-term within each sub-heading set. This can also be seen on Table II which is a term-term matrix. The reader is referred to Deerwester for the formulae used to compute Term-Term matrix and Document-Document matrix[3]. An example of second-order co-occurrence can be seen between moisture and water because of common term surface in the water heading set and in the ice heading set between Sea and Snow because of the common term, 'ice'. Term-term matrix shows a relationship between water headings and the rest of the values approach zero to indicate no relationship because no common terms exists between the two document sets. The plot shows that there is no clustering in the two document sets, the only representation shows the distance between terms in each set but not relative to the two subsets. The benefit of adding the ontology as an additional document to the corpus is to add additional semantic linkages for water related documents. The addition of the ontology illustrates that this technique can be used to cluster water related sensor resources in a web geospatial portal. To illustrate how SVD values affect term-term relationships we use two heading subsets: documents d1-d6 which are water headings and document d6-d12 which are ice headings as shown in Table I.

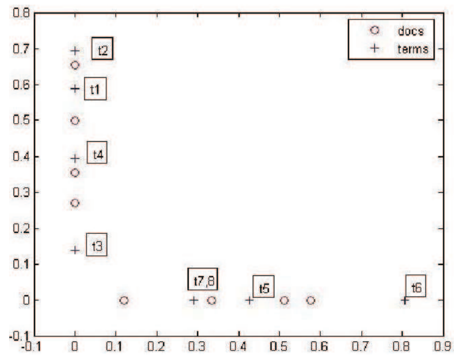

Fig. 1. Document-term plot, Water and Ice

Higher order co-occurrence is computed in LSI between

TABLE I
TERM-BY-DOCUMENT MATRIX

	Atmospheric (t1)	Water(t2)	Surface(t3)	Moisture(t4)	Sea(t5)	Ice(t6)	Reflactance(t7)	Snow(t8)
Water headings								
Atmospheric Water Droplets	1	1	0	0	0	0	0	0
Atmospheric Moisture Content	1	0	0	1	0	0	0	0
Surface Moisture	0	0	1	1	0	0	0	0
Desert Island Water	0	1	0	0	0	0	0	0
Water Usage	0	1	0	0	0	0	0	0
Surface Water Supply	0	1	1	0	0	0	0	0
Ice headings								
Sea Ice Concentration	0	0	0	0	1	1	0	0
Sea Ice Elevation	0	0	0	0	1	1	0	0
Reflectance	0	0	0	0	0	0	1	0
Ice and Snow Reflectance pixel quality assurance	0	0	0	0	0	1	1	1
Ice depth and thickness	0	0	0	0	0	1	0	0
Normalized difference snow index	0	0	0	0	0	0	0	1
Ontology headings								
Liquid water is surface water	0	2	1	0	0	0	0	0
Ice water is surface water	0	2	1	0	0	1	0	0
Gas water is atmosphere wter	1	2	0	0	0	0	0	0
Sea Water is Salt water	0	2	0	0	1	0	0	0
Sea Water is Ice Water	0	2	0	0	1	0	0	0

TABLE II
K=2 TERM MATRIX, SEPARATE WATER AND ICE HEADINGS

	Atmospheric (t1)	Water(t2)	Surface(t3)	Moisture(t4)	Sea(t5)	Ice(t6)	Reflactance(t7)	Snow(t8)
t1	0.665	1.472	0.665	0.458	0.000	0.000	0.000	0.000
t2	1.472	3.259	1.472	1.014	0.000	0.000	0.000	0.000
t3	0.665	1.472	0.665	0.458	0.000	0.000	0.000	0.000
t4	0.458	1.014	0.458	0.315	0.000	0.000	0.000	0.000
t5	0.000	0.000	0.000	0.000	1.052	1.987	0.715	0.715
t6	0.000	0.000	0.000	0.000	1.987	3.754	1.351	1.351
t7	0.000	0.000	0.000	0.000	0.715	1.351	0.486	0.486
t8	0.000	0.000	0.000	0.000	0.715	1.351	0.486	0.486

the neighbouring water terms and ice terms that appear with the term 'Depth'. The documents containing the 'Depth' term then bring the two sets together and a relationship is formed,whereas in the previous example the terms from the two document sets were disjoint. The clustering is noticably affected, by just the addition of one common term between the two term sets. Fig. 2 is a plot of the term document matrix dot product after two-dimension truncating of the original matrices. The plot shows a term document relationship as well as term-term clustering. Relationship between terms can be expressed by computing Euclidean distance between terms using cardinal coordinates from the plot. Water terms and Ice terms lie on two opposite sides of the plot while Depth which is a common term between the documents lies in the middle. The co-occurrence of Depth, forms a link between water and ice document sets. LSI computes links between documents using term co-occurrence and requires for at least one term to appear in two separate document sets in order for that link to be established. This was tested and illustrated in the previous plots where the term Depth was randomly introduced to the headings in order to create that link. Table III shows the weightings

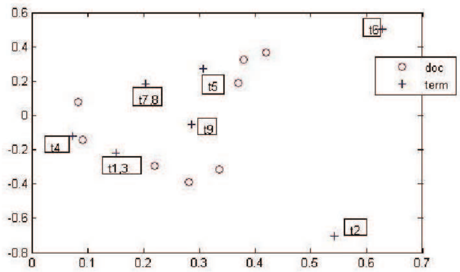

Fig. 2. Document-Term plot, joint Water and ice headings

of the term-term relationships between the two documents set after the common term has been introduced, where we see none of the weightings have zero value therefore justifying the rearranged scatter of terms from that which was seen in Fig.1.

A more semantically rich and sound approach is to create the same link using ontology concepts. In this example the Concepts used to describe water are taken from the WRSR ontology which is a sample representation of the hydrosphere concept. The idea is to create a third set of documents which contains ontology concepts. These are represented as headings which are ontological predicates. The predicates extracted from the ontology are those contained in the the original set of indexed terms. The ontology is used to link pre-existing terms. The ontology headings contain terms from both the ice headings and the water headings so by creating a set of documents with co-occurring terms, a link is established between ice documents and water documents because of the semantic relation between the two, expressed by the ontology. The predicates included in the list of ontology headings are those expressing a relationship between two indexed terms occurring on the ontology. Fig. 3 shows document clustering where three distinct clusters are formed. Ice and Water document sets have formed two separate clusters but with reduced distance between them to show a semantic link between ice concepts and water concepts. From Fig. 3a we see that effect of LSI on the term, water (t2), is to de-emphasize its effect on the clustering as it appears frequently next to different terms, this is an example of noise being removed. Second order clustering is used for the classification of water and ice headings and the ontology headings form a separate cluster where the distance between the documents and the terms show that these documents are closer to the term: water. Fig. 3a shows the term-document relationship and Fig. 3b shows the document clustering with the terms removed.

In the last plot we concentrate on the document clustering, as the previous experiments show that there is a relationship between term-occurrence and document clustering. Our interest is in the groupings of documents into the four known subheadings: d1-d6, are water headings, d7-d12 are ice headings, d13-d17 are ontology headings and d18-d22 are jewellery headings. Fig. 4 shows the plot of jewellery headings.The effect of including these headings on the previously formed clusters as well as the clustering of the headings themselves is observed to see if the semantic structure previously formed by including the ontology headings is maintained. An overlapping of terms or documents throughout the results show a strong correlation between the two, an example of this is seen on Fig. 4 between d9 and d12 which shows the strong LSI co-occurrence effect.These two documents have no common terms, but the pattern of usage of 'Reflectance' in both documents and also in the ontology documents where Reflectance is a property of ice which is snow is the cause for the correlation to be formed. The use of ontology also helps solve the synonym problem that LSI also addresses, therefore making the combination of these two methods complementary. Documents d18 and d19 are grouped with d11 which contain ice and the rest of the jewellery documents have zero coordinates. The jewellery documents are not classified along

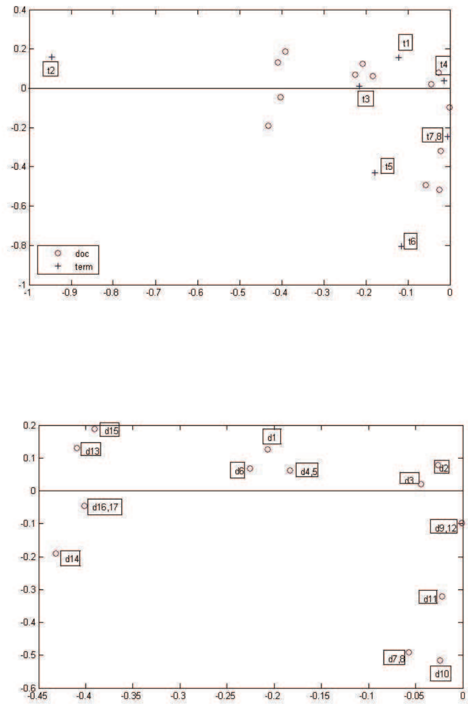

Fig. 3. Separated Document and Term plots with ontology concepts

either one of the dimensions. Note that there is no separate cluster for jewellery documents because of the bias of our ontology which only relates water and ice related phenomenon. Ontology headings have a strong relationship with each other because of the multiple occurrence of common terms, which is expected as the same superclass can have multiple subclasses and relationships with multiple other classes. The combination of ontology headings shows that because LSI reveals pattern of usage of terms, term weightings are affected by ontology headings which are representation of semantic relationships between concepts that are logically connected.

VII. Conclusion and Recommendations

The aim of this paper was to show the biasing of document clustering, using LSI, with a domain specific ontology for WRSR classification in geospatial web portals. To compare the improvement in document clustering, Ontology headings are introduced to the document corpus to create a link between documents whose extracted themes, based on indexed terms, have semantic relationships. For test purposes, a small hydrosphere concept ontology was constructed. Lastly an unrelated document set is introduced to the corpus and the results show that the previously formed document clusters are maintained

TABLE III
K=2 TERM MATRIX, JOINT WATER AND ICE HEADINGS

	Atmospheric (t1)	Water(t2)	Surface(t3)	Moisture(t4)	Sea(t5)	Ice(t6)	Reflactance(t7)	Snow(t8)	Depth (t9)
t1	0.42	1.39	0.42	0.22	-0.06	-0.06	-0.05	-0.05	0.33
t2	1.39	4.61	1.39	0.72	-0.07	0.07	-0.07	-0.07	1.17
t3	0.42	1.39	0.42	0.22	-0.06	-0.06	-0.05	-0.05	0.33
t4	0.22	0.72	0.22	0.11	-0.05	-0.06	-0.03	-0.03	0.17
t5	-0.06	-0.07	-0.06	-0.05	0.99	1.95	0.67	0.67	0.45
t6	-0.06	0.07	-0.06	-0.06	1.95	3.83	1.31	1.31	0.94
t7	-0.05	-0.07	-0.05	-0.03	0.67	1.31	0.45	0.45	0.3
t8	-0.05	-0.07	-0.05	-0.03	0.67	1.31	0.45	0.45	0.3
t9	0.33	1.17	0.33	0.17	0.45	0.94	0.3	0.3	0.52

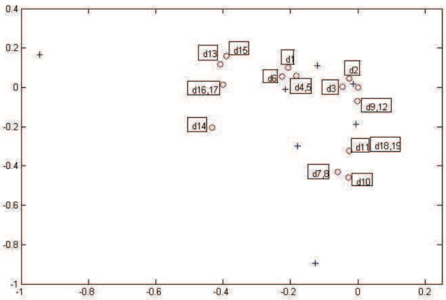

Fig. 4. Document-Term plot, joint Water and ice headings

and because of the bias of the ontology, no separate cluster is formed of the irrelevant topic, jewellery, extracted. This illustration of semantically enriched theme extraction allows for future work which entails automating the realisation process of slotting a WRSR instance under a relevant class based on document content. This automatic classification of web resident sensor resources found in portals contributes towards efforts within the Sensor Web and allows for semantic discovery of sensor systems and sensor observations filtered according to requirements that are entered by a user. Future work will also entail subclassing of ontology headings into different documents, where different weightings will be assigned to ontology concepts on different class hierarchy levels in order to semantically capture class structure, further representing structured knowledge about sensor resources within the Sensor Web.

REFERENCES

[1] T. Berners-Lee and J. H. O. Lassila, “The semantic web,” *Scientific American*, no. 284, pp. 34–43, 2001.

[2] A. A. Barfourosh, H. R. Nezhad, and D. Perlis, “Information retrieval on the world wide web and active logic: A survey and problem definition,” University of Maryland, Tech. Rep. CS-TR-4291, 2004.

[3] R. Kossala and H. Blockeel, “Web mining research : A survey,” *SIGKDD Explorations*, vol. 2(1), pp. 1–15, 2000.

[4] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, G. W. Furnas, and R. Harshman, “Indexing by latent semantic analysis,” *Journal of the American Society For Information Science*, vol. 41, pp. 391–407, 1990.

[5] P. Hitzler, R. Studer, and Y. Sure, “Description logic programs: A practical choice for the modelling of ontologies,” in *1st Wokshop on Formal Ontologies Meet Industry (FOMI’05)*, 2005.

[6] A. Kontostathis and W. M. Pottenger, “A framework for understanding lsi performance,” *Information Processing and Management*, vol. 42(1), pp. 56–73, 2006.

[7] D.Li, “Geospatial sensor web and self-adaptive earth predictive systems,” *NASA AIST PI Conference*, 2007.

[8] K. A. Delin, “The sensor web: A macro-instrument for coordinated sensing,” *Sensors 2002*, vol. 2, pp. 270–285, 2002.

[9] M. Botts, A. Robin, J. Davidson, and M. Botts, “Opengis sensor web enablement architecture document,” Open Geospatial Consortium, Tech. Rep. OGC 06-021, 2006.

[10] D. Moodley and I. Simonis, “Aa new architecture for the sensor web: the swap framework,” in *Semantic Sensor Networks Workshop ISWC’06)*, 2006.

[11] S. Chakrabarti, M. V. den Berg, and B. Dom, “Focused crawling :a new approach to topic-specific web resource discovery,” in *Proc. Eighth International Conference on The World-Wide Web*, 1999.

[12] D. Guillaume, J. Futrelle, R. McGrath, and R. Plante, “Digital library technology for locating and accessing scientific data,” *In ACM Digital Libraries ’99*, pp. 188–194, 1999.

[13] J. Heflin. (2003) Web ontology language (owl) use cases and requirements. Internet draft. [Online]. Available: http://www.w3.org/TR/webont-req/

[14] S. Melnik and S. Decker, “A layered approach to information modeling and interoperability on the web,” in *Proc. Workshop on the Semantic Web 4th European Conference on Redsearch and Advanced Technology for Digital Libraries*, 2000.

[15] D. Koller and M. Sahami, “Hierarchically classififying documents using very few words,” in *Proc. Fourteenth International Conference on Machine Learning*, 1997.

[16] G. Attardi, A. Gulli, and F. Sebastiani, “Automatic web page categorization by link and context analysis,” in *Proc. European Symposium on Telematics, Hypermedia and Artificial Intelligence THAI’99)*, 1999, pp. 105–119.

[17] M. Wang and J. Nie, “A latent semantic structure model for text classification,” in *ACM-SIGIR Workshop on mathematic/formal methods in information retrieval*, 2003.

[18] P. Zhao and L. Di, “Semantic web service based geospatial knowledge discovery,” in *Geoscience and Remote Sensing Symposium IGARSS’06*, 2006.

[19] G. Salton, A. Wong, and C. Yang, “A vector space model for automatic indexing,” in *ACM Communications*, vol. 18(11), 1975, pp. 613–620.

[20] N. F. Noy and D. McGuinness, “Ontology development 101: A guide to creating your first ontology,” Stanford KSL Technical Report, Tech. Rep. KSL-01-05, 2000.

[21] D. McGuinness, “Ontologies come of age,” in *In Spinning the Semantic Web Bringing the World Wide Web to its Full Potential*. MIT Press., 2003.

[22] D. Vallet, M. Fernandez, and P. Castells, “An ontology-based information retrieval model,” in *2nd European Semantic Web Conference ESWC*, 2005.