社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2003－ＩＣＳ－133（18）
2003／9／15

Computing Citation Relatedness Using Kernels (preliminary report)

Takahiko ITO†, Taku KUDO†, Campbell HORE†, Masashi SHIMBO†, and Yuji MATSUMOTO†

† Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
E-mail: †{takahi-i,taku-ku,campbe-h,shimbo,matsu}@is.aist-nara.ac.jp

Abstract We apply a family of diffusion kernels, in particular Chung's heat kernel and Kandola et al's von Neumann kernel, to the computaton of relatedness between technical documents based on their citation information. The advantage of using these kernels is that the diffusion process underlying its computation allows them to capture the relation between documents even when these documents do not cite or are not cited by the same documents. We compare the performance of these kernels with that of traditional co-citation and co-reference method, using real data.
Key words link analysis, bibliometrics, co-citation coupling, diffusion kernel

カーネル関数を利用した共引用分析の拡張

伊藤　敬彦†　　工藤　　拓†　　Campbell　Hore†　　新保　　仁†　　松本　裕治†

† 奈良先端科学技術大学院大学情報科学研究科
〒 630-0192 奈良県生駒市高山町 8916-5
E-mail: †{takahi-i,taku-ku,campbe-h,shimbo,matsu}@is.aist-nara.ac.jp

あらまし　技術論文の類似度を計測し手掛かりとして用いることで、必要とする論文の関連論文を検索することができる。このような情報はサーベイ活動を行なう上で大変有益である。リンク関係から論文の類似度を計測する手法として共引用分析 (同じ論文から参照される回数) と書誌結合 (同じ論文を参照する回数) が古くから使用されてきた。しかし、これらの手法では同一の論文に対して参照、被参照をしている論文同士のみに類似度が定義される。それに対し、我々は「似た論文を参照する」もしくは「似た論文から参照される」という情報を抽出する。これにより、同一の論文に対して参照（被参照）をしていなくとも適切に類似度を求めることが可能となる。具体的には、近年活発に研究が行なわれているカーネル法の中から、diffusion kernel と von Neumann kernel を使用した。
キーワード　参照情報, 共引用分析, 書誌結合, カーネル関数

1. Introduction

Recently, bibliographic references have begun to be used for analyzing technical documents. Bibliographic references indicate what other documents a given document cites or is cited by. These types of information have also been used for processing link relation among WWW pages.

There are several systems that use bibliographic information. For example, CiteSeer [12] is a system that extracts bibliographic information from scientific documents in WWW, and enable users to search for scientific documents using this information. JCR, developed by ISI, evaluates importance of technical journals on the basis of the average number of citation.

Bibliographic information has many practical use. Brandes [2] used bibliographic information for visualizing bibliographic networks. Pinski and Narin [14] proposed an algorithm for ranking journals on the basis of eigenvector, which was later enhanced by Brin and Page [3] to well-known PageRank algorithm. Kleinberg [8] also proposed an algorithm (HITS) for ranking web pages or technical documents. This algorithm ranks each document from two different perspective (authority or hub). If a document is cited by many documents with high 'hub' score, HITS algorithm gives a high 'authority' score to the document, and if a document cites many documents with high 'authority' score, it gives a high 'hub' score to the document.

Co-citation coupling [16] (a number indicating the degree to which two documents are cited by the same documents) and co-reference coupling (bibliographic coupling) [7] (a number indicating the degree to which two documents cite the same documents) are classic measures of similarity between documents using bibliographic information. Joachims et al. [6] reported that the performance of clustering web pages increased by using these similarity measures together with the contents of the page. However these coupling methods cannot measure similarity between documents when they do not cite or are not cited by the same documents

In order to solve this problem, we propose using kernel functions that are based on diffusion process to the computation of document similarity. In particular, we use Chung's diffusion (heat) kernel [4], [10] and Kandola et al.'s von Neumann kernel [9]. Kernel methods [15] have seen many successes in practical applications, and a number of kernel-compatible learning algorithms have been proposed.

In the following sections, we explain co-reference coupling and co-citation coupling (Section 2), describe our algorithms (Section 3), empirically demonstrate diffusion kernel and other kernels (Section 4), and describe of a method for integration of multiple similarity measures (Section5) .

2. Co-citation coupling and co-reference coupling

Among many types of information sources, bibliographic information is one of the most popular information sources to determine the relatedness between documents.

Co-reference coupling [7] and co-citation coupling [16] have been the most employed similarity measures using bibliographic information. Co-citation coupling is a method used to establish a subject similarity between two documents in terms of citation. And co-reference coupling is also a method establish a subject similarity between two documents in terms of reference. This information of co-citation and co-reference coupling is defined in terms of citation graph and its adjacency matrix.

Definition 1 A *citation graph* is a directed graph $G = (V, E)$, where vertices (V) represent documents and edges (E) represent citations; i.e., $(i, j) \in E$ if and only if document $i \in V$ cites document $j \in V$. The *adjacency matrix* A of a citation graph G is a matrix with its element $A_{ij} = 1$ iff $(i, j) \in E$, and 0 otherwise.

Definition 2 Given a citation graph G and its adjacency matrix A, *co-reference (coupling) matrix* and *co-citation (coupling) matrices* are defined as AA^T and A^TA, respectively. We call the weighted undirected graphs induced by A^TA the co-citation graph of G. Similarly, the co-reference graph of G is the one induced by AA^T.

An (i, j)-element of A^TA represents the value of co-citation coupling between documents i and j. Hence, $(A^TA)_{ij} = (A^TA)_{ji} = 1$, if and only if documents $i, j \in V$ are cited by the same documents. Similarly, each element of AA^T represents the value of co-reference coupling. Thus, co-citation matrix and co-reference matrix are both symmetric. Figure 3 illustrates the relation among citation graph and its co-citation and co-reference graphs.

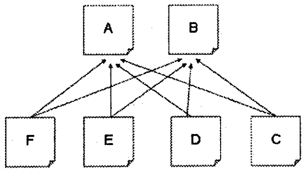


Fig. 1 co-citation coupling. A and B are similar (related) documents, because A and B are cited by same documents.

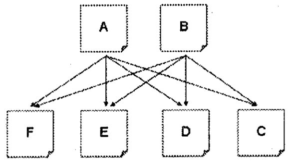


Fig. 2 co-reference coupling. A and B are similar (related) documents, because A and B cite same documents.

These methods can only compute the similarity between the documents that have a common citation or reference, or in other words documents having distance 2.

Definition 3 The *distance* between two documents in citation graph G is the length of the shortest paths in undirected graph $\tilde{G}$, where $\tilde{G}$ is obtained by ignoring the direction of the edges in G.

In Figure 4, paper A and paper C do not cite the same documents, and their distance is 4. So paper A and paper C are not related to each other in terms of co-citation coupling. However, paper B cites both paper D and paper E. This suggest that the author of paper B thought paper D and paper E are similar papers. Since B and C cite the same paper E, which as we have established is similar to D, we can infer the relationship between A and C. But neither co-citation coupling or co-reference coupling takes this aspect into account,and captures the similarity between A and C. On the contrary, our methods described in the next section compute the similarity of the documents when their distance exceeds two.

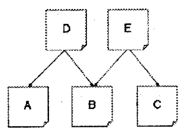
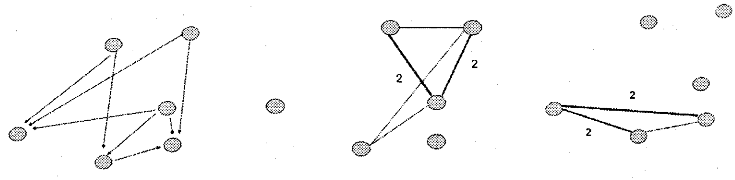


Fig. 4 The example : co-citation coupling does not work



(a) (b) (c)

Fig. 3 (a) citation graph , (b) co-reference graph, (c) co-citation graph. Both co-reference coupling and co-citation coupling can be represented as undirected weighted graph.

3. Proposed methods

Both co-reference coupling and co-citation coupling are incapable of computing similarity between papers that their distance is over two. To solve the problem, we apply *kernel methods* to citation graphs. The functions computing the inner product between mapped examples in a feature space is called kernel functions. Note that the kernel functions need not compute this inner product by explicitly mapping the examples to the feature space. Sometimes it is possible compute this inner product while retaining original representation of examples. An advantage of kernel functions is that they are compatible with many powerful machine learning methods like the Support Vector Machine [17], or clustering algorithms [1].

3.1 Diffusion kernel for computing citation similarities

The diffusion (heat) kernel is derived from spectral graph theory [4] and have been introduced to the machine learning community by Kondor and Lafferty [10]. This kernel can compute the similarity between any two given nodes efficiently without examining all possible paths between the nodes.

Definition 4 Let $G = (V, E)$ be an undirected graph, and let $n_i = |\{(i,k) \in E\}|$ be the degree of node $i \in V$. *Laplacian of* G, denoted by $L(G)$, is given by

$$L(G) = D - A$$

where D is defined by

$$D_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } (i,j) \in E, \\ n_i & \text{for } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 5 Let G be an undirected graph, and let $H = -L(G)$. The *diffusion kernel matrix* $\hat{P}(\lambda)$ on G is given by

$$\hat{P}(\lambda) = e^{\lambda H} = \sum_{t=0}^{\infty} \frac{\lambda^t H^t}{t!}$$

where λ is a decay factor.

The kernel values are given by each element of the matrix $\hat{P}$.

We can extend the diffusion kernel to weighted graphs as well, by setting each $H_{ij} (i \neq j)$ to the weight of the edge between i and j, and reweighting diagonal terms H_{ii} accordingly.

Given a citation graph G and its adjacency matrix A, it is possible to apply the diffusion kernel on co-reference matrix AA^T and co-citation matrix $A^T A$, these matrix can be used in diffusion kernel smoothly. because AA^T and AA^T represent undirected weighted graphs (see definition 2)

3.2 von Neumann kernel for computing citation similarities

Kandola [9] proposed the von Neumann kernel for the computing of document similarity based on terms. This kernel, like diffusion kernel, is also based on diffusion process, but uses Neumann series in its expansion.

Given X be documents-by-term matrix whose elements are weighted by idf (inverse document frequency) [13], document correlation $K = X^T X$ and term correlation matrix $M = XX^T$ are defined. The (i,j)-element of K gives the similarity between documents i and document j. Similarly. M gives the similarity between different terms.

Definition 6 Given a matrix X, let $K = X^T X$ and $M = XX^T$. The *von Neumann kernel matrices* $\hat{K}(\lambda)$ *and* $\hat{M}(\lambda)$ are given by the solutions of recurrences.

$$\hat{K} = \lambda X^T \hat{M} X + K \qquad \hat{M} = \lambda X^T \hat{K} X + M \tag{1}$$

where λ is a decay factor. This recurrences between $\hat{K}$ and $\hat{D}$ can be interpreted that similarity of documents are defined by terms similarities and vice versa.

Theorem 1 ([9]) If $\lambda < \|K\|^{-1}$, The solution to recursive formula (1) is given by

$$\hat{K} = K(I - \lambda K)^{-1} \quad \hat{M} = M(I - \lambda M)^{-1}$$

We now describe how to apply von Neumann kernel to citation analysis. We make two matrices to measure similarity in terms of co-reference and co-citation respectively. The first is a reference-by-documents matrix R, and each element of R are weighted by inverse document frequency. The value of (i,j)-element of R is a value dividing number of all documents by the number of documents cite j. The each value of R is called CCIDF (Common Citation Inverse times Document Frequency) [11] between documents. Without weighting, this matrix coincides with the adjacency matrix

A of citation graph G. The second is citation-by-documents matrix C, each element of C is also weighted by inverse document frequency. Without weighting, this matrix equals A^T.

To compute co-reference similarity, we use the matrix R, and make matrix $K_R = RR^T$. Without weighting, K_R would equal co-reference matrix. Then, $\hat{K_R} = K_R(I - \lambda K_R)^{-1}$ gives a similarity of documents in terms of co-reference. Similarly, To compute co-citation similarity, we use matrix C, in place of R, and matrix $K_C = CC^T$ in place of K_R.

We will later use a normalized version of these kernels. Given a kernel matrix $\hat{K}$, the normalized kernel matrix $\tilde{K}$ is given by

$$\tilde{K}_{ij} = \frac{\hat{K}_{ij}}{\sqrt{\hat{K}_{ii}\hat{K}_{jj}}}$$

The effect of normalization will be demonstrated in Section 4.

4. Experiment

The experiments in this section, analyze the performance of the kernels presented in the previous section, and compare their performance with that of other similarity measures.

4.1 Examples

We begin with a few examples that the diffusion kernel outputs to real data. The data we use is a collection containing citation information from 1682 technical documents in natural language processing. The dataset of citation was extracted by the method we [5] proposed before. We run the diffusion kernel and examined the list of documents that they considered 'similar'. In these experiments, the decay factor λ is set by 0.005.

The outputs from diffusion kernel on some documents are shown in Tables 1 and 2. The leftmost column is the ranking of similarity, and the middle column show the distance between documents.

In the above examples, diffusion kernel computed the similarity between two documents with their distance over 2. For example, in Table1, 'Bidirectional Context-Free Grammar Parsing for Natural Language Processing' and 'Semiring Parsing' have distance of 4 but they include 'parsing' in their titles, so we could guess their contents are related to each other to some extent. Similarly, 'Phrasal Translation and Query Expansion Techniques for Cross-Language Information' and 'Querying Across Languages: A Dictionary-Based Approach to Multilingual Information Retrieval' also have distance 4 but they are similar documents on the basis of their title.

Examining the results mentioned above, diffusion kernel can compute similarities between documents whose distance is more than 2.

4.2 Number of outputs

The average number of documents resulting from each method (co-reference coupling, co-citation coupling, diffusion and von Neumann kernels) were shown in Table 3. Both the diffusion kernel and von Neumann kernel assign non-zero similarity score to more documents. In our experiments, the number of outputs that diffusion kernel and von Neumann kernel return are the same, and their ranking rarely changed, even if the decay factor are changed.

Table 1 Relatedness ranking given by diffusion kernel on the co-citation similarity results on 'Bidirectional Context-Free Grammar Parsing for Natural Language Processing'

1	1	Completeness Conditions for Mixed Strategy Bidirectional Parsing
2	2	Formal Properties and Implementation of Bidirectional Charts
3	2	Tabular Method for Island-Driven Context-Free Grammar Parsing
4	2	Principles and Implementation of Deductive Parsing
5	3	Compositional Semantics for Linguistic Formalisms
6	**3**	**Semiring Parsing**
7	3	The Computational Complexity of the Correct-Prefix Property for TAGs
8	4	Compositional Model-Theoretic Semantics for Logic Programs
9	4	Modularity in Logic Programming
10	4	The Semantics of Grammar Formalisms Seen as Computer Languages

Table 2 Diffusion kernel on the co-citation similarity results on 'Phrasal Translation and Query Expansion Techniques for Cross-Language Information'

1	2	A HMM Part-of-Speech Tagger for Korean with Wordphrasal Relations
2	2	Translingual Information Retrieval: A Comparative Evaluation
3	2	Resolving Ambiguity for Cross-Language Retrieval
4	2	Query Expansion Using Local and Global Document Analysis
5	2	Word Association Norms, Mutual Information, and Lexicography
6	2	Retrieving Collocations from Text:Xtract
7	2	Aligning Sentences in Parallel Corpora
8	**4**	**Querying Across Languages: A Dictionary-Based Approach to Multilingual Information Retrieval**
9	2	The Mathematics of Statistical Machine Translation:Parameter Estimation
10	4	Extraction of Lexical Translations from Non-Aligned Corpora

4.3 Comparison to von Neumann kernel

We compare the outputs of the diffusion kernel with the von Neumann kernel on similarity of reference to examine the performance of the two kernels. In Tables 4–6, the third column is the HITS authority rank.

The outputs of von Neumann kernel of co-citation similarity are given in Table 4. The rank of 'Building a Large Annotated Corpus of English:The Penn Treebank' is 4, in the table. But we can see the document is much different from the 'Discourse Structure in Spoken Language: Studies on Speech Corpora' on the basis of their titles. 'Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text' is also the example, it seem hardly related with the target document, but the algorithm say related. Very high HITS (authority) rank are common in two documents. Using von Neumann kernel simply, The documents with hight HITS rank (authority) become very similar documents to many documents. We only want similarities between documents, so it is not good that similarity are effected by importance value.

On the contrary, the outputs of diffusion kernel (Table 5) have more similar documents than the output of von Neumann kernel and more contain the key word of target document, like 'discourse' and 'speech' than the output of von Neumann kernel. Output of diffusion kernel also are not the documents that have very high HITS rank. Thus, diffusion kernel returns more convincing output than von Neumann kernel.

Table 3 Number of outputs returned by each method

method	average
Co-reference coupling $A^T A$	19.7
Co-citation coupling AA^T	4.77
Diffusion kernel (co-reference graph)	80.4
Diffusion kernel (co-citation graph)	1124
von Newmann kernel (co-reference graph)	79.7
von Newmann kernel (co-citation graph)	1093

In order to remove the effect of HITS rank, we then use von Neumann kernel with normalization. Table 6 shows that, the ranking given by von Neumann kernel with normalization is similar to the one output of diffusion kernel.

In summary, there are two differences in the results between diffusion kernel and von Neumann kernel. The first is that we must normalize the similarity values in von Neumann kernel to remove the effect of HITS rank. On the contrary, similarity values are automatically normalized in diffusion kernel.

The second is that the similarity score of these methods. The similarity values of results of von Neumann kernel with normalization are much bigger than the that of diffusion kernel.

5. Related work

So far, in this paper, we computed separately citation similarity and reference similarity in diffusion kernel. But in practice, their information should be integrated in some tasks. In order to compute similarity among papers more precisely,

Small's similarity was used for clustering documents. This similarity measure is made by integrating four similarity measures.

Given given a adjacency matrix A of of citation graph G

- dc_{ij} (direct citation) $dc_{if} = 1$ iff $A_{ij} = 1$ or $A^T_{ji} = 1$ otherwise 0,
- bc_{ij} (co-citation coupling) $A^T A_{ij}$ or $A^T Aji$: co-citation coupling between document i and document j
- cc_{ij} (co-reference coupling) AA^T_{ij} or AA^T_{ji} : co-reference coupling between document i and document j
- lc_{ij} (longitudinal citation) lc_{ij} is the number of times i cite a paper that cites j

Given four similarity measures (dc_{ij}, bc_{ij}, cc_{ij}, lc_{ij}) between two documents i and j, Small's similarity are given by

$$s_{ij} = \frac{2dc_{ij} + bc_{ij} + cc_{ij} + lc_{ij}}{\sqrt{(1+n_i)+(1+n_j)}}$$

where n_i and n_j are degree of i and j.

Like Small's similarity, our similarity measure could be integrated. Integrating similarity measures, the relatedness among documents might be measured more precisely.

Given adjacency matrix A of of citation graph G, At fist we add each matrix that represent measures.

$$\alpha(A^T A) + \beta(AA^T) + \gamma(A + A^T) \quad (\alpha + \beta + \gamma = 1)$$

Table 4 von Neumann kernel with non normalization results on 'Discourse Structure in Spoken Language: Studies on Speech Corpora' (co-citation similarity)

1	2	87.6	33	Attention, Intentions, and the Structure of Discourse
2	4	57.3	**3**	**Building a Large Annotated Corpus of English:The Penn Treebank**
3	2	50.1	71	Assessing Agreement on Classification Tasks: The Kappa Statistic (Squibs and Discussions)
4	2	48.9	**20**	**Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text**
5	2	46.1	72	Mixed Initiative in Dialogue: An Investigation Into Discourse Segmentation
6	2	46.0	94	Centering: A Framework for Modeling the Local Coherence of Dscourse
7	2	43.8	**25**	**Multi-Paragraph Segmentation of Expository Text**
8	2	43.6	38	Empirical Studies on the Disambiguation of Cue Phrases
9	2	38.3	84	A Prosodic Analysis of Discourse Segments in Direction-Giving Monologues
10	2	36.9	106	Cues and Control in Expert-Client Dialogues

Table 5 Diffusion kernel results on 'Discourse Structure in Spoken Language: Studies on Speech Corpora' (co-citation similarity)

1	2	0.00457	266	Accent and Discourse Context: Assigning Pitch Accent in Synthetic Speech
1	2	0.00457	268	Classifying Cue Phrases in Text and Speech Using Machine Learning
1	2	0.00457	267	A Discourse Analysis Approach to Structured Speech
1	2	0.00457	270	The Intonation Structuring of Discourse
1	2	0.00457	236	Developing Algorithms for Discourse Segmentation
6	2	0.00453	257	Tense Interpretation in the Context of Narrative
6	2	0.00453	246	A Computation Theory of the Function of Clue Words in Argument Understanding
8	2	0.00440	205	Heterogenous Uncertainty Sampling for Supervised Learning
9	2	0.00437	119	Tense Tree as the "Fine Structure" of Discourse
10	2	0.00436	194	Limited Attention and Discourse Structure (Squibs and Discussions)

where $A^T A$, AA^T, and $(A + A^T)$ are represent co-citation coupling, co-reference coupling, direct link respectively. then, this matrix represent undirected weighted graph, so we can compute the diffusion process smoothly.

6. Conclusions

We proposed new methods for measuring documents similarity on the basis of kernel methods, a diffusion kernel and von Neumann kernel. Our methods compute similarity of citation and reference respectively,like co-citation coupling and co-reference coupling. In addition our methods can compute the similarity between documents when their distance is over 2.

We tested the performance of our methods with examples in real data, and showed our methods can return the related documents to target document, even if they do not cite or are not cited by the same documents. Then we compared the performance between diffusion kernel and von Neumann kernel. Finally, we proposed integration

Table 6 von Neumann kernel with normalization results on 'Discourse Structure in Spoken Language: Studies on Speech Corpora' (co-citation similarity)

1	2	0.551	268	Classifying Cue Phrases in Text and Speech Using Machine Learning
1	2	0.551	236	Developing Algorithms for Discourse Segmentation
1	2	0.551	267	A Discourse Analysis Approach to Structured Speech
1	2	0.551	270	The Intonation Structuring of Discourse
1	2	0.551	266	Accent and Discourse Context: Assigning Pitch Accent in Synthetic Speech
6	2	0.434	119	Tense Tree as the "Fine Structure" of Discourse
7	2	0.430	196	Evaluating Automated and Manual Acquisition of Anaphora Resolution Strategies
8	2	0.376	205	Heterogenous Uncertainty Sampling for Supervised Learning
9	2	0.372	246	A Computation Theory of the Function of Clue Words in Argument Understanding
10	2	0.367	103	Intention-Based Segmentation: Human Reliability and Correlation with Linguistic Cues

of our methods, on the basis of Small's similarity.

7. Acknowledgment

We thank Daichi Mochihashi for valuable comments and discussions.

References

[1] A. Ben-Hur, H. T. Siegelmann, and V. Vapnik. Support vector clustering. In *Journal of Machine Learning Research*, Vol. 2, pp. 125–137, 2001.

[2] U. Brandes and T. Willhalm. Visualization of bibliographic networks with a reshaped landscape metaphor. 2002.

[3] S. Brin and L. Page. The anaomy of a larg-scale hypertextual(web)search engine. In *In The Seventh Inernational WWW Conference*, 1998.

[4] F. R. K. Chung. *Spectral Graph Theory:CBMS Regional Conference Series in Mathematics, No 92*. American Mathematical Society, 1997.

[5] T. Ito, S. Horibe, M. Shimbo, and Y. Matsumoto. Citation indexing using many similarity measures. In *IPSJ SIG Technical Reports 2003-DBS-130*, pp. 181–188, 2003.

[6] T. Joachims, N. Cristianini, and J. Shawe-Taylor. Composite kernels for hypertext categorisation. In *In Proceedings of the International Confeence of Machine Learning(ICML)*, 2001.

[7] M. M. Kessler. Bibliographic coupling between scientific papers. In *American Documentation Vol.14,No.1*, pp. 10–25, 1963.

[8] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. In *Journal of the ACM*, 1999.

[9] J. Kondola, J. Shawe-Taylor, and N. Cristianini. Learning semantic similarity. 2001.

[10] R. Kondor and J. Lafferty. Diffusion kernels on graphs and ohter discrete input spaces. In *In Proceedings of the International Confeence of Machine Learning(ICML)*, pp. 21–24, 2001.

[11] S. Lawrence, K. Bollacker, and C. L. Giles. Autonomous citation matching. In O. Etzioni ed., *Proceedings of the Third International Conference on Autonomous Agents*, New York, 1999. ACM Press.

[12] S. Lawrence, C. L. Giles, and K. Bollacker. Digital libraries and autonomous citation indexing. *IEEE Computer*, 32(6):67–71, 1999.

[13] C. D. Manning and H. Schutze. *Foundations of Statistical Natural Language Processing*. MIT Press, fourth edition, 1999.

[14] G. Pinski and F. Narin. Citation influence for journal aggregates of scientific publications: theory, with appplication to the literature of physics. In *Information Processing and Management*, Vol. 12, pp. 297–312, 1976.

[15] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2002.

[16] H. Small. Co-ciation in the scientific literature:a new measure of the relationship between two documents. In *Journal of American sociiety for information science Vol.24*, pp. 265–269, 1973.

[17] V. N. Vapnik. *The Nature of Statistical Learning Threory*. Springer, 1995.