

# CAUSAL KNOWLEDGE EXTRACTION BY NATURAL LANGUAGE PROCESSING IN MATERIAL SCIENCE: A CASE STUDY IN CHEMICAL VAPOR DEPOSITION

Yuya Kajikawa\*, Yoshihide Sugiyama, Hideki Mima, and Katsumori Matsushima

\*Institute of Engineering Innovation, School of Engineering, the University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, 113-8565, Japan

Email: [kaji@biz-model.t.u-tokyo.ac.jp](mailto:kaji@biz-model.t.u-tokyo.ac.jp)

## ABSTRACT

*Scientific publications written in natural language still play a central role as our knowledge source. However, due to the flood of publications, the literature survey process has become a highly time-consuming and tangled process, especially for novices of the discipline. Therefore, tools supporting the literature-survey process may help the individual scientist to explore new useful domains. Natural language processing (NLP) is expected as one of the promising techniques to retrieve, abstract, and extract knowledge. In this contribution, NLP is firstly applied to the literature of chemical vapor deposition (CVD), which is a sub-discipline of materials science and is a complex and interdisciplinary field of research involving chemists, physicists, engineers, and materials scientists. Causal knowledge extraction from the literature is demonstrated using NLP.*

**Keywords:** Natural language processing, Knowledge extraction, Materials science, Chemical vapor deposition, Causality

## 1 INTRODUCTION

Science has expanded its knowledge by the subdivision of a research domain and specialization of its sub-domain. A vast array of knowledge produced in subdivided and specialized domains has been accumulated mainly in journal articles. If we assume the subdivision rate and specialization rate are constant, we can reasonably speculate that the number of scientific domains and resulting scientific knowledge grows exponentially with time as already pointed out (de Solla Price, 1963). However, the individual scientist has to focus on or specialize in only a few scientific subdomains to keep up with the growth of the domains.

Although specialization is an inevitable strategy, in order to obtain deeper scientific understanding, it is common sense that the flood of information and specialization make it difficult for scholars to obtain a comprehensive perspective not only on research domains different from their own domain, but also on their own specific research domain (Ziman, 2001; Yamaguchi & Komiyama, 2001; Börner, Chen & Boyack, 2003; van der Vet, 2000). Ziman (2001) concisely describes this situation in his article discussing scientific

education as, "Research scientists are trained to produce specialized bricks of knowledge, but not to look at the whole building." It is a natural consequence that such a loss of comprehensive perspective degrades the productivity of an individual work. We sometimes feel it difficult to grasp the whole perspective even on a topic we are familiar with or to catch up with current research and even a previously known result. It is also a daily problem that when we look through recent scientific papers, we find one reporting result already known or easily inferred from the previous studies (Kajikawa, Noda & Komiyama, 2003; Kajikawa, Noda, & Komiyama, 2004). A typical example is the crash of the H-II 8 rocket in Japan in 1999. One of the reasons for the crash is attributed to the deficiency of knowledge; the mechanical designer did not have knowledge of materials science. He used fatigue property data measured by NASA, but the actual material in Japan has a smaller grain size and therefore a weaker fatigue property (NASDA, 2003). The relation between grain size and fatigue property is fundamental for a materials scientist but not for a mechanical designer. Such a situation was typically shown by an interesting but rather ethically problematic experiment performed by Peters & Ceci (1982).

They picked out 12 articles already published in 12 prestigious psychological journals and resubmitted these papers to the same journals while changing the authors' names, changing affiliations into fictional ones, and cosmetic modification of the title, abstract, and introduction. Consequently, only 3 out of 12 articles were detected as re-submission by 3 out of 38 editors and referees. Nine articles were sent for peer review. Only one article was accepted for publication, and 8 articles were rejected. Peters and Ceci focused on the result that only one article was accepted among the 12 articles published and stated as a summary of this experiment that the most important factors deciding the fate of submitted articles were the authors' status and institutional affiliation. They proposed the introduction of a blind author system for referring articles to eliminate the problem.

However, another and more important point indicated by this experiment is the fact that only 3 referees could detect the re-submissions. This indicates that neither editors nor referees have sufficient knowledge even on a specific research domain. It must also be pointed out that this can also be true not only for reviewers, but also for all of us, individual researchers. It seems that scholars are currently required to keep up with the pace of information growth whose speed is ever increasing. Traditional approaches are brute-force in nature - researchers have to sort through mountains of literature to conduct their surveys. Obviously, this is time-consuming, difficult to repeat, and subjective. To overcome the situation, we need research support systems to retrieve, access, and extract knowledge from the previous literature.

Our aim of study is to develop further a support for research activity using natural language processing (NLP). Currently, we focus on knowledge extraction by NLP, which is applied to the journal articles of materials science and engineering (MSE). The aim of this paper is to overview the previous efforts to support research activities and to introduce our system. A case study in chemical vapor deposition (CVD), which is a sub-discipline of materials science, is shown.

## **2 BACKGROUND**

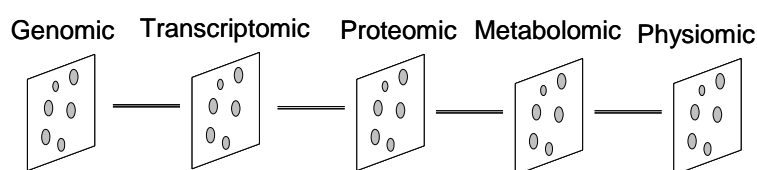
Computational support of scientific activity is indispensable for scholars in the present day. Traditionally, knowledge has been stored in libraries as books, articles, and encyclopedias. Nowadays, the repository is changing into e-libraries, which assist us in retrieving and accessing the relevant literature. The aim of this paper is to enhance the function of e-libraries by NLP. NLP is a term representing the process dealing with text written in natural language by computer. Typical examples of NLP in applications are information retrieval and machine translation. The NLP technique improves the effectiveness of information retrieval in e-libraries by using thesaurus, document sorting, and document clustering. Knowledge extraction is another promising candidate to improve the usability of e-libraries.

Knowledge extraction by NLP consists of the following steps: term recognition, synonym resolution, and knowledge extraction. The first step, term recognition, extracts keywords among the corpus. Usually, the corpus is handled by statistical technique, and the importance of each term is judged by statistical indicators such as token frequency and document frequency. Currently, more sophisticated indicators for term recognition are being developed. Examples are C-value and C/NC-value (Mima & Ananiadou, 2000). The second step is synonym resolution. In this step, terms expressing the same concept are united. The final step is knowledge extraction. In this step, the relations among collected concepts are analyzed. The most used technique is co-occurrence of concepts. The basic hypothesis underlying this is that the co-occurrence of concepts in the same document or sentence is the track of some relations among those concepts. NLP is expected, as an efficient and feasible technique, to extract plausible causal relationships in large corpus and to enforce scholars to handle a large quantity of scientific literature.

The most developing discipline of knowledge extraction by NLP in application is bioinformatics because of its importance, larger knowledge accumulation, and production than those in other disciplines. The accumulated knowledge in biology is apparently larger than those that each researcher can manipulate. NLP in bioinformatics is utilized to extract information such as protein, gene, and small molecule interactions from scientific literature (Ono, Hishigaki, Tanigami, & Takagi, 2001; Yandell & Majoros, 2002; Rzhetsky, Iossifov, Koike, Krauthammer, Kra, Morris et al. 2004). Knowledge extraction by NLP is established as a cross-disciplinary area between biology and computer science. Currently, computer science oriented conferences such as the Text Retrieval Conference (TREC) have a section on genomics and biology, and conferences related to biology such as the Pacific Symposium on Biocomputing have a section on NLP and text mining.

These studies are motivated by the pioneering work of Swanson (Swanson, 1986). Swanson proposes that combining existing, though unconnected, knowledge results in new knowledge: one publication stated the relationship between the two phenomena A and B, while another reported on the relationship between the phenomena B and C; if no one has reported on the association between A and C, this association can be considered to be new knowledge. The crucial notion in this view is that two pieces of information are not related directly: there is only a hidden connection.

Swanson's discovery concerns Raynaud's disease (Swanson 1986). Raynaud's disease causes intermittent blood flow in the extremities (fingers, toes, and ears). There was neither a general treatment nor cure for Raynaud's disease. Swanson surveyed the literature and formulated the hypothesis that fish oil might be used for treating Raynaud's disease. Studying the literature on Raynaud's disease (i.e., C), Swanson observed that many blood- and blood vessel-related characteristics (i.e., B) are typical for Raynaud's patients. Swanson also found that fish oil (i.e., A) lowered blood viscosity and platelet aggregation. However, there was no article relating fish oil with Raynaud's disease. Later, Swanson's discoveries were corroborated experimentally and clinically (Smalheiser & Swanson 1998). This induced researchers to mimic Swanson's discovery using NLP.



**Figure 1.** Knowledge structure in molecular biology. Dots and quadrangles represent concepts and omic spaces, respectively.

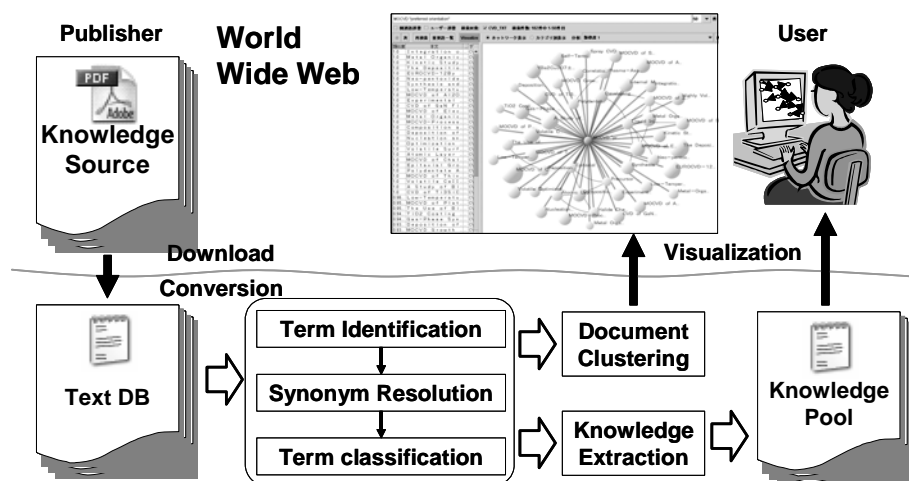
Generally, knowledge in molecular biology can be represented by relations among concepts consisting of genomic, transcriptomic, proteomic, metabolomic, and phenomic space (Toyoda & Wada, 2004). One simple method to identify the interactions among these concepts is to extract words corresponding to concepts in each space co-occurring in the same article or, more reliably, in the same sentence. Such co-words may show a kind of relation among them, though it is often noisy (Gordon & Lindsay, 1996; Weeber, Klein, de Jong-van den Berg, & Vos, 2001). Therefore, a straightforward method is to extract words co-occurring in the same sentence with some rules. For example, Weeber et al. (2001) extracted the cure for a disease, in this case, Raynaud's disease, by using a semantic filter. They extracted concepts relating to Raynaud's disease from 1,246 articles. Those articles included 317,205 tokens as single word, bigrams, or trigrams. All these tokens are potential candidates for a cure for Raynaud's disease. They excluded meaningless terms by using the Unified Medical Language System (UMLS) which is a semantic medical dictionary (Lindberg, Humphreys & McCray, 1993). They obtained 5998 concepts registered in UMLS among 317,205 tokens. They narrowed down the 5998 concepts into 145 concepts using a semantic filter such as focusing on concepts having biological function or cell function. Results extracted by this method showed definite higher precision than those extracted by only statistical methods such as token frequency and record frequency.

We can use other clues to improve the accuracy of NLP. Ono et al. (2001) proposed a method to extract protein-protein interaction in scientific literature by pattern matching. They focused on 4 selected verbs: interact, associate, bind, and complex. By focusing on such sentences where protein A interacts with protein B, protein-protein interactions are extracted with a high recall rate (>80%) and precision rate (>90%). By using this parsing scheme, a biological database with huge amounts of data is constructed.

Compared to the considerable number of studies in biomedicine, unfortunately, to our knowledge, there is no study on applying NLP for MSE. This motivated us to test the validity of NLP to extract knowledge from MSE literature. We conducted a case study of knowledge extraction from CVD literature.

### 3 CASE STUDY

CVD is a technology depositing solid materials from the gas phase via chemical reaction and is now becoming an indispensable technology that is suited to various applications such as coatings, particle production, and microelectronic devices. The preparation of high purity metals, insulators, and semiconductors has developed significantly in the last 50 years or so. These progresses are motivated by the significant demands and requirements of the microelectronics industries. Currently, over 100,000 pieces of literature related to CVD including journal articles, patents, and conference proceedings are reported. As noted by Hitchman (1995), CVD is the complex and interdisciplinary field of research involving chemists, physicists, engineers, and materials scientists. That literature is traditionally scattered throughout a wide range of journals, from physics, applied physics, organic and inorganic chemistry, application, and MSE. The journal, *Chemical Vapor Deposition*, was launched in 1995 as a forum for specialists engaging in CVD. Currently, the journal is recognized as one of the leading journals among the CVD community and is ranked as one of the journals with the highest impact factor in the ISI category, Materials Science, Coatings & Films. The goal of our research is to analyze all available MSE literature. However, in this paper, we only focus on this specific journal as a case study. We analyzed the full texts of the literature published in *Chemical Vapor Deposition* from 1998-2003.



**Figure 2.** Overview of system and NLP processing.

Figure 2 is an overview of our system. We downloaded 257 articles including about 750,000 words. Those files were converted into text files. Term identification was performed using the C-value method (Mima & Ananiadou, 2000). C-value indicates the relative importance of sequential strings in the corpus, which is given when a is nested as,

$$C - value(a) = \log_2 |a| \cdot f(a) \tag{1}$$

otherwise,

$$C - value(a) = \log_2 |a| \left\{ f(a) - \sum_{b \in T_a} f(b) / P(T_a) \right\} \tag{2}$$

where  $a$ ,  $|a|$ ,  $f(a)$ ,  $T_a$ ,  $P(T_a)$  are the candidate string, the length of  $a$ , its frequency of occurrence in the corpus, the set of extracted candidate terms that contain  $a$ , and the number of these candidate terms, respectively. In short, the C-value has a high value when the term with long strings frequently appears in the corpus. The extracted terms by the C-value method are shown in Table 1.

Table 1 List of terms extracted by C-value method.

Rank	Term	C-value	Rank	Term	C-value
1	gas phase	505.7	11	300 C	177.5
2	deposition rate	483.7	12	500 C	175.9
3	growth rate	419.5	13	Table 1	164.6
4	deposition temperature	312.1	14	other hand	161.7
5	thin films	277.8	15	film thickness	157.6
6	room temperature	232.2	16	X-ray diffraction	154.8
7	400 C	220.3	17	350 C	137.9
8	substrate temperature	211.4	18	thermal decomposition	137.5
9	carrier gas	208.1	19	flow rate	136.5
10	residence time	181.5	20	Figure 1	135.1

As can be seen, the list has noisy terms such as 400 C, Table 1 etc. We manually exclude these terms by registering them as so-called stop words. While it is not possible to identify keywords only by the C-value method, it can help us to extract keywords. After listing the keywords, synonyms were manually assembled into the same concept. This is a necessary step because the same concept is often expressed by a variety of terms. For example, the concept superconducting transition temperature is expressed by terms such as superconducting transition temperature, superconducting critical temperature, critical temperature, and  $T_c$ . Finally, we extract causal knowledge among concepts by using the co-occurrence of concepts in the same sentence.

Table 2. List of concepts co-occurred with superconducting transition temperature.

Rank	Term	Times	Rank	Term	Times
1	carbon	10		magnetic susceptibility	2
2	niobium nitride	9		nickel	2
3	Substrate temperature	7		resistance	2
4	annealing	6		carbon impurity	2
	superconducting critical current density	6		superconductivity	2
6	YBCO	4		vacancy	2
7	Auger electron spectroscopy	3		Ba concentration	2
8	c-parameter	3		Ba/Y ratio in vapor	2
	lattice parameter	3	19	barium titanate	1
10	Curie temperature	2		buffer layer	1

Table 2 is a list of concepts co-occurring with the superconducting transition temperature. Ninety-eight sentences and 49 concepts co-occurred with the superconducting transition temperature. They are ranked by the number of times of co-occurring sentences. These concepts resemble some types of relationships but not all of them are causal relationships. We counted the relevance of the extracted sentence. Twenty-eight sentences among 98 sentences show causal relationships between the concept and superconducting transition temperature. Therefore, the precision rate is 28.6%. The results derived from simple co-occurrence of those concepts were very noisy as expected. These noises causing a low precision rate are mainly by errors in the identification of concepts and relationships other than causality. The example of the former is that  $T_c$  is identified as concepts of both *superconducting transition temperature* and *Curie temperature*. The example of the latter is that YBCO film deposited at 700°C has  $T_c$  of 40 K. This ‘has-a’ relationship is a typical example of a non-causal relationship. The former is difficult to overcome when literature is written in natural language, but in the latter case it is possible to raise the precision rate.

As discussed in Section 2, one of the methods to improve the accuracy is to use verbs. However, the methodology is highly domain specific. For example, 4 verbs used by Ono et al., i.e., interact, associate, bind and complex, seem irrelevant to us to express the knowledge in MSE. And we cannot find any other expression, partly due to the small set of corpus. Another method to improve the accuracy is to use semantic filtering, which sets meaningful relations at the meta-level in advance (Weeber, 2001). For example, in molecular biology, meta-knowledge can be represented by relations among concepts consisting of genomic, transcriptomic, proteomic, metabolomic, and phenomic space as shown in Fig. 1. Therefore, by only focusing on the relationship, knowledge is effectively extracted with high precision. In order to apply semantic filtering to MSE literature, we must formulate the meta-knowledge of MSE.

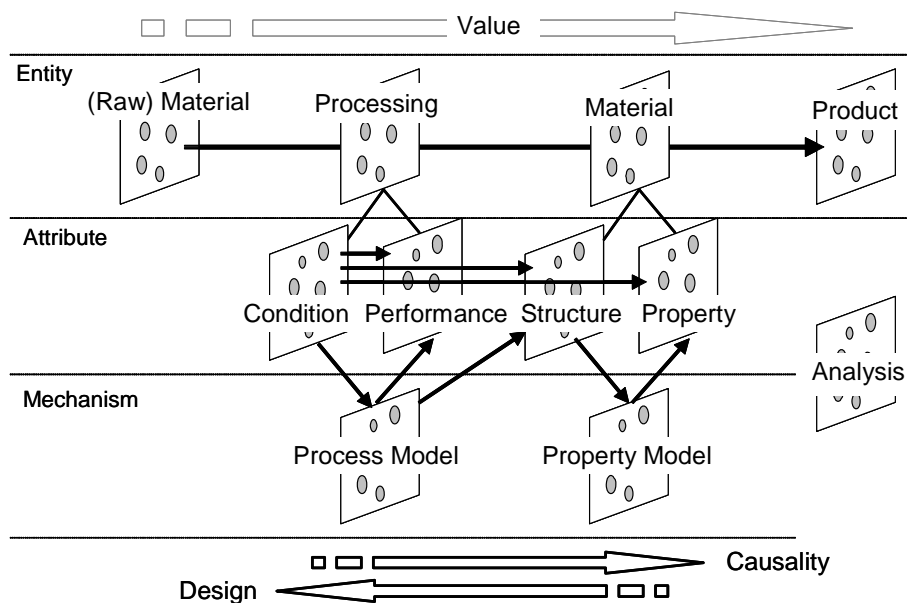


Figure 3. Knowledge structure in MSE.

In MSE, it is generally accepted that knowledge is a set of relations among processing, structure, property, and performance. An example is the relation between grain size and fatigue property as already shown in the rocket example, which is the causal relation between structure and property. This view on the knowledge structure of MSE, processing, structure, property, and performance, reaches a general agreement while some detailed discrepancy exists (White, 1999, and Yamaguchi & Komiyama, 2000). However, the above four categories focusing on the attribute of entities do not cover the concepts appearing in the current corpus. In addition to attribute line, there is a mechanism line underlying each relationship between attributes, analysis to monitor the attribute, and an entity line which is the transforming process of materials in the real world. Therefore, we classified concepts into 10 categories based on the above view as shown in Fig. 3. The 10 categories are *material* (such as silicon, SiH<sub>4</sub>), *processing* (CVD, anneal), operating *condition* (substrate temperature, total pressure), process *performance* (deposition rate, yield), materials *structure* (surface roughness, carbon content), materials *property* (electric resistivity, refractive index), *analysis* (X-ray diffraction, scanning electron microscopy), *process model* (surface diffusion, grain growth), *property model* (conduction, absorption), and *product* (thin film transistor). The goal of MSE is to understand the causal relation between attributes (condition, performance, structure, and property) in order to control and design *performance* and *property* by adjusting the *condition* and *structure*. To do so, the underlying mechanism (*process model* and *property model*) should be elucidated.

Based on this classification schema, semantic filtering was performed to extract causal knowledge. For example, assuming only the relation expressed by an arrow in Fig. 4 is causal knowledge, plausible candidates that control superconducting transition temperature are narrowed down to 17 concepts. Nineteen sentences among 30 sentences show the causal relationships between the concept and superconducting transition temperature. Therefore, the precision rate in the case of semantic filtering is 63.3%, which is higher than 28.6% by simple co-occurrence. By using semantic filtering, we can extract the meaningful relations by retaining a relatively high precision rate. The results of our experiments on the superconducting transition temperature are summarized in Fig. 4. Such a diagram helps the researchers to efficiently extract knowledge among the corpus.

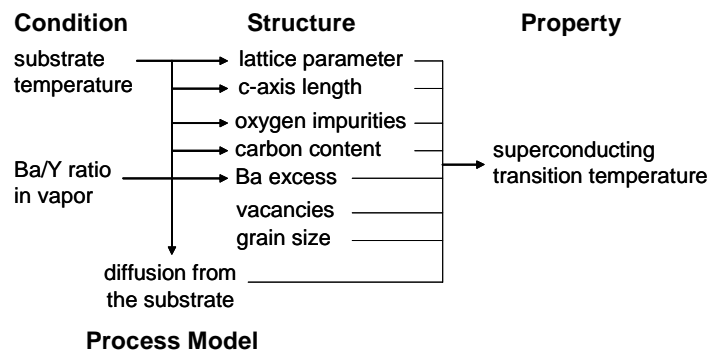


Figure 4: Path diagram affecting superconducting transition temperature extracted by semantic filtering.

Currently, we are gathering these sentences reporting causal knowledge in the knowledge pool (Fig. 2). We



are also conducting document clustering to enable rapid sorting of documents. Combining these tools, knowledge extraction and document clustering, we can efficiently handle a large number of documents and extract knowledge written there. To validate the performance of our system quantitatively, the recall rate and precision rate should be validated toward a large experimental set in large corpus. Our current validation is carried out only in small corpus and experimental sets; however, it shows that how NLP can be used in MSE and semantic filtering using simple classification schema greatly improves the accuracy in the extraction of causal knowledge.

## **4 CONCLUSION**

Due to the specialization of scientific research and information flood, currently, it is extremely difficult to obtain a comprehensive perspective. Therefore, to promptly retrieve relevant documents and extract knowledge from them, a research support system is demanded. Natural language processing (NLP) is one of the promising candidates and is being widely studied in the biomedical discipline. However, there is no work in materials science and engineering (MSE). In this study, we constructed a knowledge extraction system.

At present, researchers seldom take enough time to read journal articles. It is reported that your average natural scientists spend slightly less than two hours per week reading journals, and 60% of this "reading" involves scanning titles and abstracts of journals as they arrive on the scientist's desk (Cole, 2000). Considering the current publish or perish syndrome, it is not an exaggeration to say that each scientist writes but does not read journal articles. Our system can be utilized for the fast sorting and reading of a large number of documents within a limited time.

However, knowledge extraction by NLP still has a problem in each processing step. In term recognition, eliminating noisy terms and adding terms uncovered by NLP are still laborious processes despite the help of the NLP tool. Use of nomenclature is also a problem in synonym resolution. For example, researchers often described  $T_c$  as the abbreviation of the superconducting transition temperature. However, it is identical to the atomic symbol of technetium, although fortunately current corpus does not include technetium. Another is complexity of the process name. Process is often named as a combination of processing and condition. For example, many process names related to CVD exist such as atomic pressure (AP) CVD, low pressure (LP) CVD, plasma-enhanced (PE) CVD, *etc.* This type of nomenclature seems to unnecessarily increase the complexity of technical terms and reduce the effectiveness of NLP. In order to enhance the usability of e-libraries to the utmost extent, action for standardization or reconsideration of publication style in scientific literature is necessary. However, standardization inevitably involves costs. The combination of the current style of literature, i.e., written in natural language, and NLP is not perfect but works to some extent. Therefore, it may still continue. In this case, further work of NLP to improve the efficiency and effectiveness is still required.

## 5 REFERENCES

Börner, K., Chen, C., & Boyack, K.W., (2003) Visualizing knowledge domains. *Ann. Rev. Info. Sci. Tech.* 37, 179-255.

Cole, S. (2000) The role of journals in the growth of scientific knowledge. In Cronin, B. and Atkins, H. B., (Eds.), *The Web of knowledge: A Festschrift in Honor of Eugene Garfield*, Medford, New Jersey, Information Today Inc.

de Solla Price, D.J. (1963), *Little science, big science*, New York, Columbia University Press.

Gordon M.D. & Lindsay, R.K., (1996) Toward discovery support system: A replication, re-examination, and extraction of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil. *J. Am. Soc. Info. Sci.* 47(2), 116-128.

Hitchman, M. (1995) A professional forum for the field of chemical vapor deposition. *Chem. Vap. Deposition*, 1(1), 5-7.

Kajikawa, Y., Noda, S., & Komiyama, H., (2003) Comprehensive perspective on the mechanism of preferred orientation in reactive-sputter-deposited nitrides. *J. Vac. Sci. Technol. A* 21(6), 1943-1954.

Kajikawa, Y., Noda, S., & Komiyama, H., (2004) A simple index to restrain abnormal protrusions in films fabricated using CVD under diffusion-limited conditions. *Chem. Vap. Deposition*, 10(4) 221-228.

Lindberg, D.A.B., Humphreys, B.L., & McCray, A.T., (1993) The unified medical language system. *Methods of Information in Medicine*, 32, 281-291.

Mima, H. & Ananiadou, S. (2000) An application and evaluation of the C/NC-value approach for the automatic term recognition of multi-ward units in Japanese. *Terminology* 6(2) 175-194.

NASDA (2003) Retrieved Nov. 31, 2004 from the National Space Development Agency of Japan (NASDA) web site: [http://www.nasda.go.jp/press/2003/02/datasheet-sac\\_20030226\\_j.html](http://www.nasda.go.jp/press/2003/02/datasheet-sac_20030226_j.html)

Ono, T. Hishigaki, H. Tanigami, A. & Takagi, T. (2001) Automated extraction of information on protein-protein interactions from biological literature, *Bioinformatics* 17(2), 15-161.

Peters, D.J. & Ceci, S.J., (1982) Peer-review practices of psychological journals – the fate of accepted published articles, *Behav. Brain Sci.* 5(2), 187-195.

Rzhetsky, A., Iossifov, I., Koike, T., Krauthammer, M., Kra, P., Morris, M., Yu, H., Dubou, P.A., Weng, W.,

Wilbur, W. J., Hatzivassiloglou, V., & Friedman, C. (2004) GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data, *J. Biomed. Info.* 37, 43-53.

Smalheiser N. R. & Swanson D.R., (1998) Using ARROWSMITH: a computer-assisted approach to formulating and assessing scientific hypotheses. *Computer Methods and Programs in Biomedicine* 57, 149-153.

Swanson, D. R. (1986) Fish oil, Raynauds syndrome, and undiscovered public knowledge, *Perspectives in Biology and Medicine* 30, 7-18.

Toyoda T. & Wada, A. (2004) Omic space: coordinate-based integration and analysis of genomic phenomic interactions. *Bioinformatics* 20(11), 1759-1765.

van der Vet, P.E. (2000) Building web resources for natural scientists. *Interactive distributed multimedia systems and telecommunication services (Proceedings IDMS2000)* (pp. 205-210). Berlin.

Weeber, M., Klein, H., de Jong-van den Berg, L.T.W., & Vos, R., (2001) Using concepts in literature-based discovery: Simulating Swanson's Raynaud-fish oil and migraine-magnesium discoveries. *J. Am. Soc. Info. Sci. Technol.* 52(7), 548-557.

White, M. A. (1999), *Properties of Materials*, New York, Oxford Univ. Press.

Yamaguchi, Y. & Komiyama, H. (2001) Structuring knowledge project in nanotechnology materials program launched in Japan. *J. Nanoparticle Research*, 3, 105-110.

Yandell, M.D. & Majoros, W. H., (2002) Genomics and natural language processing, *Nat. Rev. Genet.* 3(8), 601-610.

Ziman, J., (2001) Getting scientists to think about what they are doing. *Science and Engineering Ethics* 7(2), 165-176.