2011 International Conference on Emerging Intelligent Data and Web Technologies

PoLyInfo: Polymer Database for polymeric materials design

Shingo Otsuka*, Isao Kuwajima†, Junko Hosoya†, Yibin Xu† and Masayoshi Yamazaki†
*Department of Information and Computer Sciences,
Kanagawa Institute of Technology, Kanagawa, Japan
Email: otsuka@ic.kanagawa-it.ac.jp
†Materials Database Station,
National Institute for Materials Science (NIMS), Ibaraki, Japan
Email: kuwajima.isao, hosoya.junko, xu.yibin, yamazaki.masayoshi}@nims.go.jp

***Abstract*—Polymer database (PoLyInfo) provides various data required for polymeric materials design. The main data source is academic papers on polymers. PoLyInfo collects information on polymer names, chemical structures, processing methods of samples, measurement conditions, properties, used monomers and polymerization methods. PoLyInfo continues the expansion of data as one of NIMS materials database. This paper presents system architecture of "Polymer database (PoLyInfo)" and its distinctiveness. And we also report usage survey of our system.**

***Keywords*-materials database; polymer database; chemical structure search; polymer search**

I. Introduction

We have opened NIMS materials database (MatNavi)[1] in the Internet since April, 2003[2] as materials researchers and apparatus designers obtain the information about a materials. MatNavi is comprised of following things and we place it with the portal site about materials.

- Data of basic properties about materials as atom, molecules, crystal structure.
- Data of foundational engineering.
- Search system and applications using database.

We show a webpage of MatNavi in Figure 1 and main databases in Table I. In this paper, we describe system architecture, search method and usage survey of polymer database (PoLyInfo) which has many entry data and users in the MatNavi.

There are various problems for the construction of the polymer database in consideration of materials design support. First, the contents of entry data come in a variety of types. Moreover, it is necessary to design the database that can systematically collect polymer information because the polymer is composed of a lot of molecules, and there is a characteristic with complex relativity between molecules. For example, there are a view of materials and a view of compound of ethylene (monomer) in polyethylene that is typical polymer. Therefore, it is necessary to store information of them in the databases. The structure of the compound registry system (polymer dictionary) for the polymer is necessary to store polymer as compound in databases accurately. Hence, it is possible to identify chemical structure and to accommodate fluctuation of description of trivial name.

As results of grappling with these problems, 'PoLyInfo' that is the polymer database system has been opened to the public since April, 2001[3]. This system has the data model of the polymer and the function of the compound registry system for the polymer. It goes continuing enhancing data and the system now.

It is the first attempt to discuss cyclopaedically the system architecture, the search strategy and the usage survey in PoLyInfo. Though about nine years have passed since PoLyInfo was open to the public and there were some reports concerning development concept and system survey in PoLyInfo[5], [4], [6], [8], [10], [3].

There is some large-scale database that relates to the polymer. For example, National Institute of Advanced Industrial Science and Technology (AIST) provides spectrum database of organic compound although there is not a lot of data concerning the polymer. And Korea Research Institute of Chemical Technology (KRICT) provide 'Chemical materials Information Bank' service. However we obtain the data of the materials made business in this system only because the information is extracted from catalog data of the South Korea enterprise. There is materials information service of "MatWeb" in the United States. This service provides a lot of data such as ceramics and the metals besides the polymer. However it is a data only of the materials made business offer because it uses the catalog data of the enterprise as well as the example of KRICT. Besides this, there is service of "Polymers: A Property Database" that the publisher of the scientific journal offers. This service has aimed to make the user inspect the papers (pay contents). And the use fee of "SciFinder" that is the article retrieval service of American Chemical Society is high. Hence the users can not use the system airily.

In this way, these services provide useful information concerning the materials but there are some problems (amount of data is low, use fee is high, etc.). The advantage of our system (PoLyInfo) is free access to a lot of polymer data extracted from articles.

II. System architecture of polymer database (PoLyInfo)

PoLyInfo is composed of the simulation part (properties estimation) and the database part as shown in Figure 2.

[1]http://mits.nims.go.jp/index_en.html
[2]NIMS is National Institute for Materials Science.
[3]NIMS manages this system now though Japan Science and Technology Agency (JST) manages in March, 2003.

978-0-7695-4456-4/11 $26.00 © 2011 IEEE
DOI 10.1109/EIDWT.2011.13

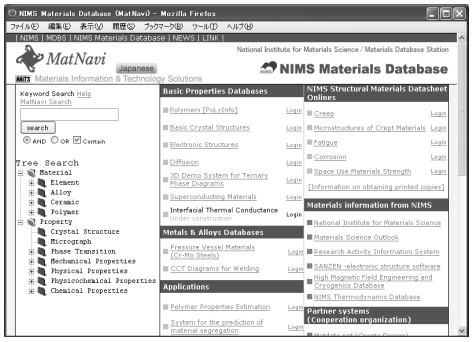

Figure 1. Webpage of "NIMS Materials Database".

Table I
NIMS MATERIALS DATABASES (MATNAVI).

No.	Database Name
1	Database for Electronic Structures
2	Basic Database for Crystal Structures [Pauling File]
3	3D Demo System for Ternary Phase Diagrams
4	Diffusion Database
5	Superconducting Materials Database
6	Polymer Database [PoLyInfo]
7	NIMS Structural Materials Datasheet Onlines (Creep, Fatigue, Corrosion, Space Use Materials Strength)
8	Microstructures of Crept Materials
9	CCT Diagrams for Welding
10	Database System for Pressure Vessel Materials
11	Nuclear Materials Database
12	Nuclear Reaction Database for Materials
13	Thermophysical Property Prediction System for Composites [Compo Therm]

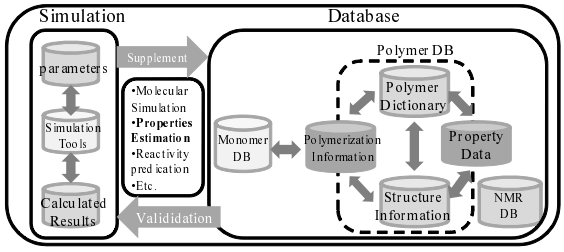

Figure 2. Structure of PoLyInfo.

In the simulation part, an application estimate properties of data not registered and virtual materials using real data in database. In the properties estimation, there is a group contribution method in one of the techniques for properties estimation from the chemical structure of constitutional repeating unit (CRU) of polymer. We use properties estimation of Van Krevelen[9] in PoLyInfo and properties of prediction object are shown as a relational expression of some factors. It calculates as sum of contribution from the atom group (atomic group parameter) included constitutional repeating unit (CRU). This method is possible to use database correlate with chemical structure and properties value with effect because the relational expression and the atomic group parameter are computable from the analytical result of the real data.

In the database part, we divide polymer information into properties data, structure information and polymerization information and polymerization information cooperates with the monomer database as the users do not have to consider data model though it is internally maintained. About information of name and structural unit, it is possible to treat polymer which is different structure with a same composition type using polymer dictionary. We call polymer registered to PoLyInfo *polymer sample* and we employ hierarchical data structure to express one polymer sample efficiently.

The information about each polymer sample as the primary structure information that is registered in polymer dictionary and forming method is correspondence to every polymer sample. And detail information about the real polymerization is included in the polymerization information.

III. A STORAGE METHOD OF POLYMER DATA

In PoLyInfo, at first, we analyzed the polymer data using OMT method(OMT:Object Modeling Technique)[7], built the data model and stored ObjectStore for build the polymer database in consideration of materials design support. After that time, we rebuilt this system using PostgreSQL because object-relational database was diffuse.

A. Analysis of the polymer information

Representative polymer information includes polymer name[4], Polymer classification[5], polymer properties[6]. Moreover, Solid state properties of polymer differ from stereoregularity, polymerization degree, forming method and measurement method even as the same polymer name. Hence, the system take in a lot of information about the monomer, the polymer chain, the polymer aggregate and the polymer materials to treat the characteristic of polymer sample (the experiment sample) cyclopaedically.

The monomer information includes a name, a molecular formula, a structural drawing. The polymer information include a structural unit (name, chemical structure, molecular weight), the primary structure (stereoregularity, molecular weight). And there is the polymerization information that is information to link the polymer chain to the monomer. The polymer aggregate include information of higher-order structure (information of polymer processing and crystal architecture) and properties information. Moreover, the polymer materials have materials name and commodity name. Therefore, PoLyInfo is possible to analyze above information for building data model.

B. Collecting of polymer information

The data source of PoLyInfo is the papers which registered Chemical Abstracts Service (CAS). We select papers to register with PoLyInfo among the following view points.

[4]International Union of Pure and Applied Chemistry(IUPAC)-compliant name of structural foundation, a name of materials foundationCa popular nameCabbreviated expressions and so on.

[5]A polymer has the classifications such as a homopolymer, a copolymer, a polymer blend. On account of the space, we omit it about the detailed contents.

[6]thermal properties, electric properties, mechanical properties and so on

Table II
SCIENTIFIC JOURNALS WITH MANY PAPERS REGISTERED WITH POLYINFO.

No.	Journal name	Number of paper
1	Journal of Polymer Science	2,745
2	Macromolecules	2,455
3	Journal of Applied Polymer Science	2,230
4	Polymer	1,893
5	Makromolekulare Chemie	620
6	Polymer Preprints (American Chemical Society, Division of Polymer Chemistry)	609
7	Synthetic Metals	602
8	European Polymer Journal	541
9	Polymer Engineering and Science	421
10	Annual Technical Conference - Society of Plastics Engineers	418
11	Journal of Macromolecular Science	370
12	Polymer Journal	322
13	Journal of Chemical Physics	319
14	Journal of Physical Chemistry	309
15	Macromolecular Chemistry and Physics	290
16	Langmuir	281
17	Polymer International	275
18	Journal of the American Chemical Society	269
19	Journal of Applied Physics	251
20	Proceedings of SPIE	251

Table III
NUMBER OF OPEN DATA IN POLYINFO.

Homopolymers	10,270
Copolymers	3,364
Polymer Blends	1,097
Composites	1,505
Monomers	16,390
Property points	220,189
Literature data	12,843

– The chemical structure is clear.
– The actual value includes various properties.
– The method for measurement and measurement condition are listed.

We also register the calculation value using actual measurement value (derived figure) and do not register citatory value.

The specialists in polymer do screening for 2,800 CAS papers per year manually and select polymer sample from 700 papers per year. They write properties information in data sheet and register with PoLyInfo (actually PostgreSQL).

The scientific journal with many papers registered with PoLyInfo is shown in table II. And brief overview of registry data is shown in table III.

One polymer sample is comprised of data sheets to have five hierarchies where a top is paper information in Figure 3. As for the polymer data, there is difference in primary structure, the highly advanced structure about the same polymer class. And the value of properties data change by a difference of method for measurement and the measurement condition in the same polymer sample. Hence, PoLyinfo

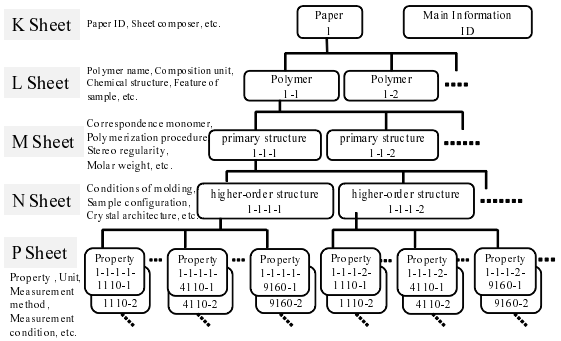

Figure 3. Outline of Data Sheet.

Table IV
CATEGORY OF PROPERTIES.

Physical property	Flexural property
Optical property	Compression characteristic
Thermal property	Creep characteristic
Electrical property	Heat characteristic
Physicochemical property	Impact strength
Dilute solution property	Hardness
Rheological property	Heat resistance and Combustion
Tensile property	Other property
Shear property	

adopt hierarchical data structure because the scientific journal for the collection is described the properties of one or more polymer samples.

K sheet in figure 3 are described basic information as the sheet composer and paper ID that polymer samples are listed in[7]. We make a plural L, M and N sheet depending on the number of the polymer and the structure if plural polymer samples are listed in a paper[8]. Moreover, the physicality value of each polymer sample is registered with the P sheet. And we make a plural P sheet like other seats if the measurement method and its condition of experiment are plural in the same polymer sample. The P sheet is divided into plural seats depending on properties because the properties information that the sheet composer can register for one polymer sample is more than 100 items as thermal properties, electric properties and mechanical properties in Table IV.

C. Polymer Dictionary

For method of registration polymer information, we manage polymer data using a compound registry system for polymer which we call it *polymer dictionary* and we made. We assort and identify polymer using chemical structure of

[7]We manage the bibliography information such as the author name of papers, the name of journal (magazine) at the other table.

[8]The information about the polymer is registered with the L sheet. And the detail information of the name and the structure are registered with polymer dictionary.

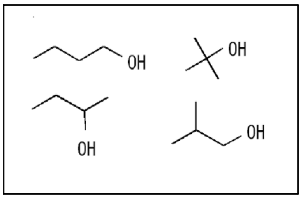

Figure 4. Isomers of $C_4H_{10}O$.

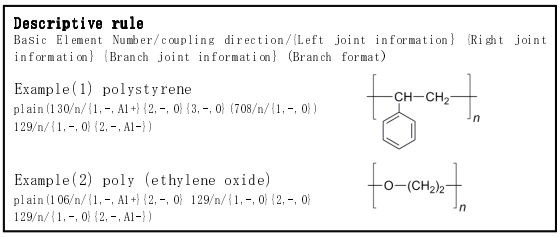

Figure 5. Structural Format of PoLyInfo.

constitutional unit in polymer dictionary. The materials of constitutional formula are different even if a composition type is the same. In general, chemical physical properties are different. We call such a thing a constitution isomer. For example, Figure 4 shows a part of the structural isomer of $C_4H_{10}O$[9]. The structure is different by a place with $-OH$ though all compositional formulas are same[10]. The chemical structure of the constitutional unit express in combination of atomic group which constitutes the chemical structure of the constitutional unit. The polymer name is based on IUPAC nomenclature system and sheet composer give a structural foundation name and a raw materials foundation name[1], [2]. A structure format of PoLyInfo is shown in Figure 5.

The format that is recorded in the polymer dictionary becomes uniformization by the identification algorithm of polymer dictionary. And the polymer dictionary has compositional formula with the structural unit base, chemical formula weight and the polymer name like structural foundation name in others of the chemical structure information. Therefore, it is possible to search based on various names and chemical structures in polymer search feature to describe section IV.

Early versions of PoLyInfo treat homopolymer only. Now, PoLyInfo treat also co-polymer and polymer blend by the expansion of the recent polymer dictionary.

Moreover, PoLyInfo can manage blend dictionary, monomer dictionary, polymerization dictionary and the data of journal information in others of the polymer dictionary. Herewith, it is possible to manage various information uniquely.

[9]We show the example of the monomer daringly to facilitate understanding.

[10]We call upper left on the Figure 4 *n-butyl alcohol*, left lower *sec-butyl alcohol*, upper right *tert-butyl alcohol* and right lower *isobutyl alcohol*

D. Data structure for polymer search

A database to use in PoLyInfo consists of data sheet information and polymer dictionary mentioned in previous section, journal information and indexes for efficient search.

The data of the L, M and N sheet are comprised of a table of one each, as shown in Figure 6. Each table has the information about the characteristic, sheet ID and polymer ID. And P sheet is divided into about 60 tables every properties as P3110, P7110. P sheet also has properties information, sheet ID and polymer ID like other tables.

The sheet ID succeeds to the sheet ID of the high rank hierarchy. First, paper ID becomes the K sheet ID. Next, the L sheet ID added serial numbers of L to the K sheet ID. In a similar way, M, N and P sheet ID added serial numbers of it to the sheet ID of the upper level. We call ID of N sheet in Figure 6 *sample ID* (for example, 29429-16-1-1 and 00081-1-1), we use this ID for the presentation such as lists of search results in PoLyInfo.

The polymer dictionary has polymer master table which is comprised of polymer name, polymer group, structural data and compositional formula and monomer master table which is described monomer information. Moreover, it has polymerization path table to manage the linkage of the polymer and the monomer and blend master table which is comprised of plural polymer. For example, ID 75227-3 of L sheet in Figure 6 is comprised of plural polymer and blend ID is given. It is possible to know sample information of this blend to search blend master table using this blend ID. Polymer master table has polymer ID and there is correspondence relation to polymer ID that is maintained on each seat.

In addition, there is the table that paper information recording the polymer sample in PoLyInfo and this table include the author information. PoLyInfo has index table consisting of sample ID, polymer ID and the minimum and the maximums of each properties for the efficient of the polymer sample search using the value of properties.

E. The advantage of using hierarchical data structure

The search for the upper level sheet is possible to delete the most right value of current sheet ID and to do exact match search because each sheet ID includes the ID of the upper level sheet. The search for the lower level sheet is possible to do prefix search using current sheet ID. It can anticipate a speedup of the search time because prefix search processing is fast in generically. For example, if the users want to know correspondence monomer and molecular weight of sample ID *00081-1-1-1* (N sheet ID) in Figure 6, the users search ID *00081-1-1* in M sheet table. And the search about the properties of the polymer sample should search ID to begin with *00081-1-1-1* for each P sheet. For example, the users understand that the polymer ID of *00081-1-1 -1* is *P01002* in N sheet table and get polymer information of compositional formula and name to search polymer master tables because each sheet has polymer ID. About the search of properties as rolling temperature of the glass and electric permittivity, the users get sample ID quickly using index tables.

Polymer search uses polymer master table at first. Polymer master table has structural foundation name, raw materials

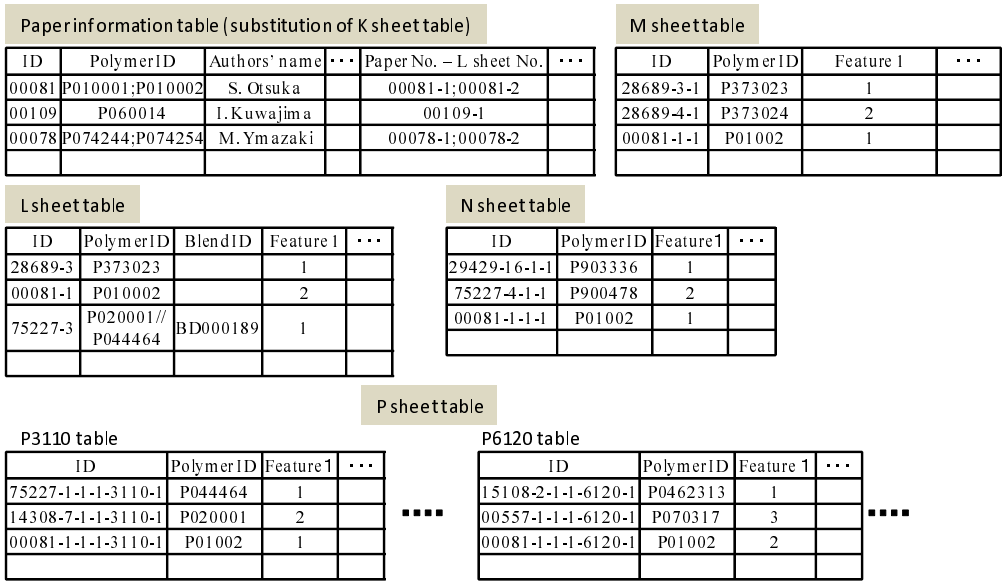

Figure 6. Example of Data Sheet Tables.

foundation name, popular name and structural data. The users can get polymer ID corresponding to these names in performing a text search for these columns. After, the users can get the information about the polymer sample corresponding to these names. For monomer search, the users get monomer ID using monomer master table and get monomer ID to search master table of polymerization path using the monomer ID. Polymer search, polymer structural search and monomer search are realized by searching these tables.

The implementation of PoLyInfo avoid join operator that is high load processing in SQL because the data structure of PoLyInfo is hierarchic. The search time is slow even if making index for searches to each ID column because each ID is resisted as character string. And it may take time by the presentation of search results because a case becoming the partial search occurs in polymer and monomer search. Now, the enrollment of the polymer master table is around 20,000 and we replaced a database system from ObjectStore to PostgreSQL. When users perform a complicated search, results are displayed in one or two minutes. Search time may increase in future when amount of data growth. We recognize that it is necessary to improve it in this connection.

The sheet composers create the data sheet and the polymer dictionary using Microsoft Excel. The update work of the database and the rebuilding work of the index table are batch processing[11].

IV. Polymer search feature

The search feature of PoLyInfo includes three kinds (polymer search, polymer structural search and monomer search).

[11]These works are around several times in a year.

A. Polymer search

The users can two type of search in PoLyInfo. One is basic search and the other is advanced search. The screenshot of basic search is shown in Figure 7. Search items of basic search are polymer type, polymer name, polymer classification, materials type[12], polymer properties and paper information.

For basic search of polymer name, the users can choice an element by select menu other than text search. For basic search of polymer properties, the users can decide the range of the value of properties. For basic paper search, the users can specify journal name, author's name and publication year. It is possible to 'ANDf search basically when the users choice plural items.

For advanced search, the users can specify various fields as average molecular weight, sample configuration, addition agent and crystallinity degree. And the number of properties items is three kinds.

B. Polymer structural search

Polymer structural search is search feature to specify subject to the chemical structure of the structural. The method of structural search are two way. One is easy structural search and the other is advanced structural search. The users search can specify the kind and the number of the atomic group of the structural using more than three hundred base element in the easy structural search. The advanced structural search can specify bonding state of chemical structure using Java applet. The screenshot of easy structural search is shown in Figure 8 and advanced structural search is shown in Figure 9.

[12]In addition to this, there are polymer, composite and compound. On account of the space, we omit it about the detailed contents.

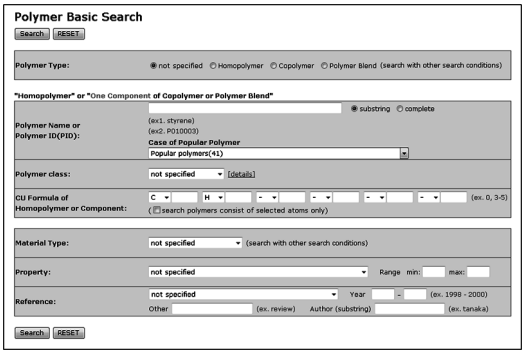

Figure 7. Easy Plymer Search.

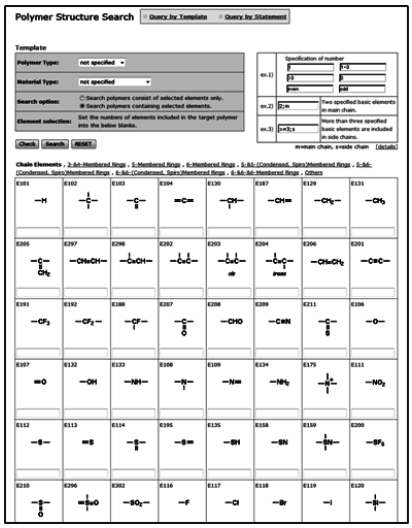

Figure 8. Polymer Structural Search using Atomic Selection.

C. Monomer search

The monomer search which is polymer raw materials is possible to search to specify monomer name, registry number of various chemical compound, molecular formula and molar weight.

D. Search result and transition of usersf behavior

The users can next search and show details easily using current search results in PoLyInfo. Transition of user behavior after search is shown in Figure 10. First PoLyInfo displays the number of polymer, properties item and measurement value each structural unit which matched search condition (Polymer list of in Figure 10) if the users do polymer search. Next, it displays *Sample list* listing the polymer of the structural unit. Finally, it displays *Sample information* displaying all information including the properties of each polymers. The users can check crystal condition in *Histogram* which can move from *polymer list*.

And the users can show *Constitutional Unit information (CU Information)* to move from *polymer list*. The CU Information displays names of the structural unit, compositional

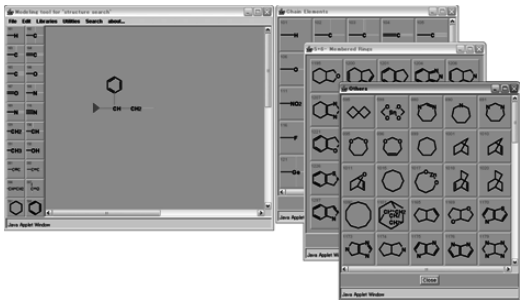

Figure 9. Polymer Structural Search using Original Tools.

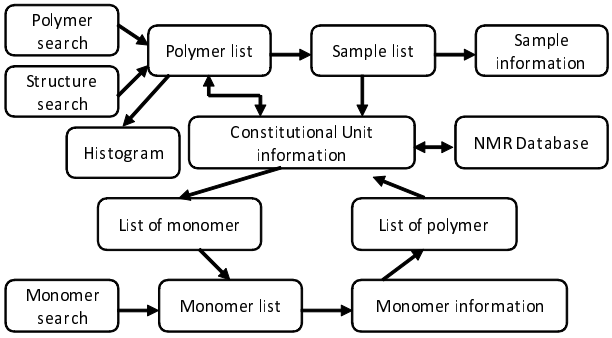

Figure 10. Transition of User Behavior.

formula and chemical structure and the users can move to *List of monomer* to show the monomer of candidate raw materials and polymerization classification, *co-polymer* which assumes the polymer concerned a structural unit, polymer blend and NMR(Nuclear Magnetic Resonance) information. [13]. In the monomer search, *monomer information* displaying all information of the monomer is displayed via *Monomer list* to show results list.

V. Usage of our system

A. Access trend of MatNavi

First, we describe access trend of MatNavi which is the portal site about a materials database including PoLyInfo. User's registration is necessary for using each database in MatNavi. The users can use all databases including the basic crystal structures database, diffusion database and so on as well as PoLyInfo by performing user's registration once. There are 53,850 registered users, 141 countries and 16,730 organizations at March, 2011. So MatNavi is one of the most famous materials database in the world. Figure 11 shows an institute (organization) and area with much registration users. The access trend of MatNavi is shown in Figure 12 and there are around one million accesses every month. And the access trend of a main database is shown in Figure 13. The Polymer database is the most access in MatNavi[14].

[13] We get bonding state of atomic in detail using NMR.

[14] In Figure 12 and 13, the access means the number of the hits (the simplest count method including the access of images files).

Figure 11. Users Locations.

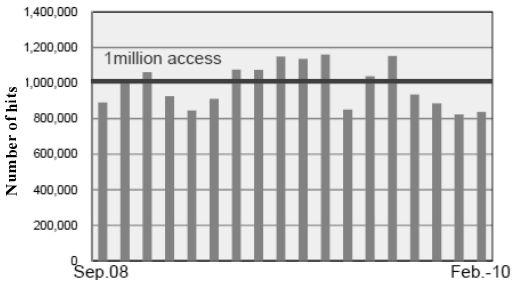

Figure 12. Transition of a Number of Hits for MatNavi.

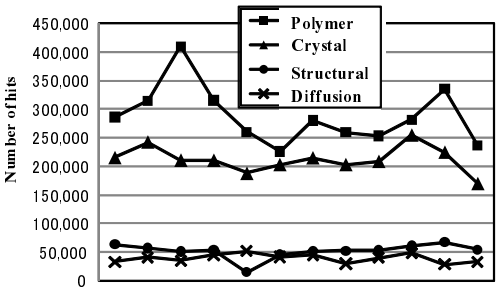

Figure 13. Transition of a Number of Hits for Main Databases.

B. Study of access trend

We use *Apache* for a Web server of MatNavi[15]. Web access logs are divided into every database because we install the Web server in every each database[16]. We analyze web access logs from April, 2006 to March, 2009.

At first, we unified all web access logs as a preliminary arrangement to analyze the logs. And we cleaned it using following steps.

– Delete image files and css files.
– Delete web crawler access.
– Delete access to error pages.

Transition of a number of accesses for MatNavi included all database is shown in Figure 14. The number of the access

[15]We remove local access from access logs.
[16]Afterward, we assume that the databases have web server function

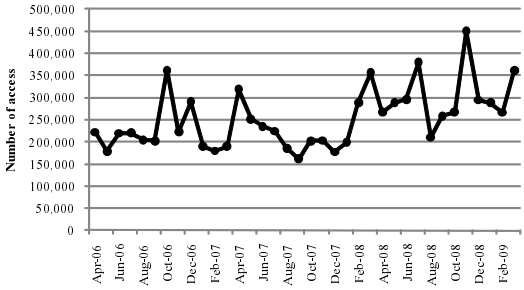

Figure 14. Transition of a Number of Accesses for MatNavi.

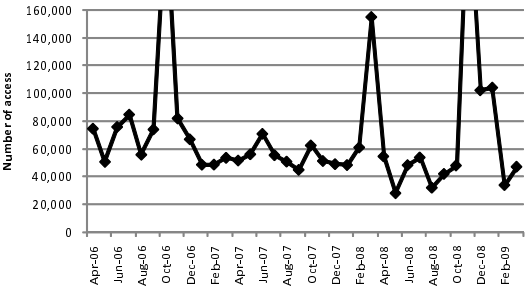

Figure 15. Transition of a Number of Accesses for PoLyInfo.

increases year by year although the number of the access spikes by each month. The number of the access in August is low because many users have a vacation.

Transition of a number of accesses for PoLyInfo is shown in Figure 15. The average of number of the access is 40-50 thousands per month although the number of the access spikes by each month. The number of the access at October, 2006 is high in order to attack by the automatic collection program.

C. A difference of the database use situation by the user

We analyze the use situation of database in MatNavi by a Japanese university and the Japanese company because web access logs include access IP information. The term of analysis is from April, 2008 to march, 2009 and we added up the number of the access each day. The results of PoLyInfo access is shown in Figure 16 and the results of basic crystal structures database (Pauling file) is shown in Figure 17. As results, the number of access from Japanese university in basic crystal structures database is high. And the number of access from Japanese company in PoLyInfo is high. We know that the Japanese company performs a polymer research lively.

VI. Conclusion and Future Work

In this paper, we describe system structure, search method and usage of PoLyInfo (and MatNavi).

Now, it is possible to treat homo-polymer, co-polymer and polymer blend in PoLyInfo. However, it is not possible to treat branched polymer like dendrimer and hyperbranched

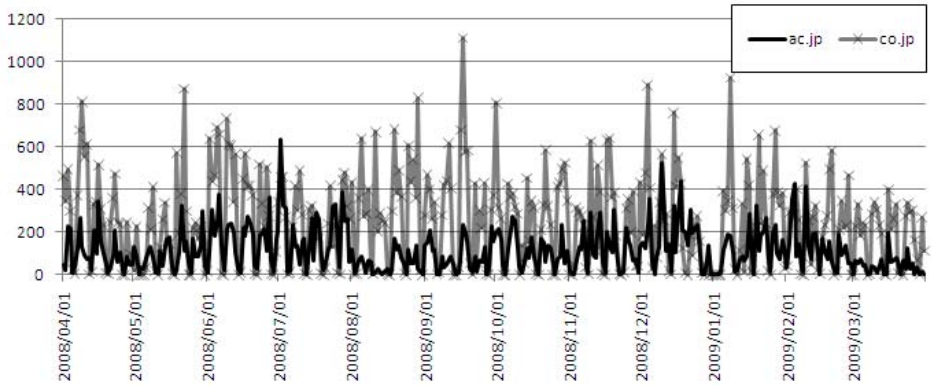

Figure 16. Difference of Access Trend between Universities and Companies (PoLyInfo).

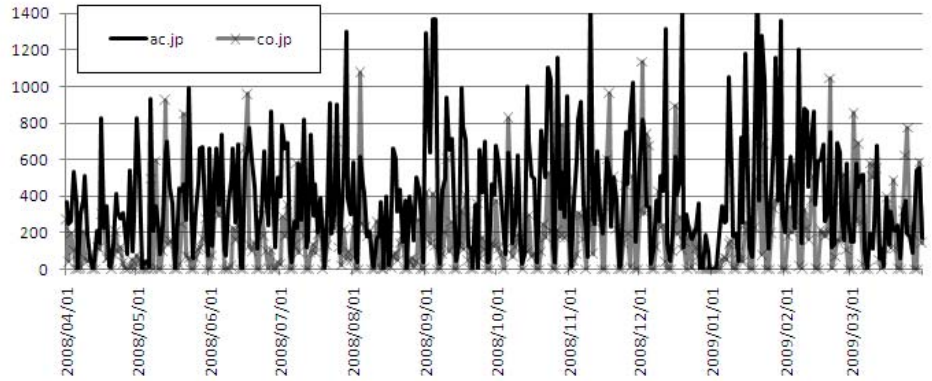

Figure 17. Difference of Access Trend between Universities and Companies (Basic crystal Structures Database).

polymer. We level up the polymer dictionary for collect and express all polymers that structure is clear.

And the work to collect data from papers and to make each dictionary is carried out with human power. From the viewpoint of budget, simplification of this work is one of the important problems. In the future, we modify our system about growth of properties, speedup of search and user-friendly interface.

ACKNOWLEDGMENTS

This work was supported by a Grant-in-Aid for Young Scientists (B) from Japan Society for the Promotion of Science (JSPS) under Grant No. 21700124.

REFERENCES

[1] P. A. Chem. Nomenclature of regular single-strand organic polymers. *Rules Approved 1975*, 48:373–385, 1976.

[2] J. Committee of Polymer Nomenclature, The Society of Polymer Science. *Terminology and Nomenclature of Polymer*. Kodansha, 2007.

[3] I. Kuwajima, M. Kumazaki, K. Mashita, J. Hosoya, and M. Yamazaki. Current status and future development of ”polymer database (polyinfo)”. In *The 2nd Asian Materials Database Symposium (AMDS 2010)*, pages 3E–01, March 2010.

[4] T. Maeda. Construction and use of polymer database. *POLYMERS*, 48(5):320–323, 1999.

[5] T. Maeda. Role of database to polymer research. *POLYMERS*, 51(6):442–453, 2002.

[6] K. Nagasaka. Construction of polymer physical properties database and physical properties prediction approaches. *CICSJ Bulletin*, 22(3):38–43, 2004.

[7] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorensen. *Object-Oriented Modeling and Design*. Prentice Hall, 1991.

[8] K. Shiono, T. Maeda, J. Kawashima, S. Iimuro, K. Mashita, K. Nagasaka, Y. Takaeda, Y. Yoshinari, and K. Matsuda. Polymer database ”polyinfo”. *JCPE Journal*, 13(4):225–234, 2001.

[9] D. van Krevelen. *Properties of Polymers, 3rd edition*. Elsevier, 1990.

[10] M. Yamazaki. Current status of polymer database (polyinfo). *CICSJ Bulletin*, 25(4):103–107, 2007.