

# HOKKAIDO UNIVERSITY

Title	Design and Implementation for Scientific Article Data Base
Author(s)	Saito, Tatsuki
Citation	北海道大學工學部研究報告 = Bulletin of the Faculty of Engineering, Hokkaido University, 151: 19-34
Issue Date	1990-07-30
Doc URL	<a href="http://hdl.handle.net/2115/42240">http://hdl.handle.net/2115/42240</a>
Right	
Type	bulletin (article)
Additional Information	

Instructions for use

## Design and Implementation for Scientific Article Data Base

Tatsuki SAITO\*

(Received March 31, 1990)

### Abstract

Scientific article information data base system ANGEL was designed based on the data base management system ADABAS. It consists of an interactive input and retrieval system, intelligent input system and a system for the generation of a relation matrix. These systems were implemented by NATURAL of ADABAS. ANGEL is available for other scientific article data base systems. Initially, ANGEL was developed as an article data base relating to the sculptural surface generation theory of CAD (Computer Aided Design)/CAM (Computer Aided Manufacturing).

ANGEL is now open for public usage and is used at Hokkaido University Computer Center. In this paper, the design concept of the system and input system of bibliographic items are described, and several results obtained from the system are discussed. The input object of ANGEL is scientific articles written in English.

### 1. Introduction

Considering valid use of a scientific article data base, as well as use of bibliographic information, it is important to use relational information among articles effectually in order to reproduce higher intellectual information, and it is the motivation of this paper. Thereupon, the necessity of a system to comprehend a research trend was examined and it was investigated what information makes the essence of research clear. Attention was paid to the various relations among scientific articles. The relations are expressed by a labeled-graph, in which an article is represented as a point and a relation between articles is represented as a line. The aim of this paper is to develop the methodology for analyzing characteristics of the graph.

A citation relation among articles is an important information of relations. Accordingly, it is inputted with information of cited place because it expresses the significance of citation.

A keyword relation also is one of the necessary information to represent the commonness of the content or the theme. If there are many common keywords between two articles, then these articles are interpreted to be treated with a similar theme or the same

---

\* Department of Precision Engineering, Faculty of Engineering, Hokkaido University, Sapporo 060, Japan

theme. These information among articles are the objective relations of information that are not influenced by the person to examine.

Relational graphs of articles are made out from these various information using this data base system, and some relational matrices are created from these graphs. The matrices are similarity matrices which have real values. After setting a mathematical criterion according to a purpose of processing, they are unified into one similarity matrix finally in order to clarify the property of the relational structure by cluster analysis. Thus configuration of the system is the target of this paper. In consequence, the system is designed to attach importance to an effectual input of these information.

The data base system ANGEL consists of keyboard subsystem, bibliographic item extraction subsystem, retrieval subsystem, display subsystem and storage subsystem in the entire composition. There are some shared modules which are entities of subsystems. These subsystems were implemented in NATURAL on ADABAS.

## 2. Constitution of the Data Base ANGEL

The constitution of the data base ANGEL is shown in Fig. 1. Bibliographic information of scientific articles is put into the data base system by manual operation directly or by optical character reader automatically through recording media like a floppy disk device. The most outside frame (thick lined) means the data base management system ADABAS that is the relational model type for main frame computer. The chained lined frame means the data base system ANGEL which consists of four parts. The first part (upper-left) is the interactive input and retrieval system, and the second part (upper-right) is the intelligent input system. The third part (middle) is the data storage (file). The fourth part (lower) is the generating system of relation matrix among scientific articles in the data base. The interactive input and retrieval system, and the intelligent input system are constructed respectively by four subsystems, that is, keyboard subsystem, display subsystem, retrieval subsystem and storage subsystem. The keyboard subsystem of the interactive input and retrieval system is for interactive manual input or interactive retrieval from VDT (visual display terminal).

On the other hand, the bibliographic item extraction subsystem of the intelligent input system is for automatic input from a recorded file. The display subsystem is for displaying processed results in the interactive input and retrieval system or in the intelligent input system. In the intelligent input system, the processed results are displayed in order to verify them.

The retrieval subsystem is to check whether processing data overlaps recorded data or not, and the subsystem of the interactive input and retrieval system enables to retrieve by author name, title, keyword et cetera. The storage subsystem is to store not only bibliographic data but also the logging data of processing date, processing type (store, update, delete), user-ID, the last article number or the last article-ID to be accessed. An example of the logging data is shown in Fig. 2.

The generating system of the relation is to set up the initial relation matrix from a

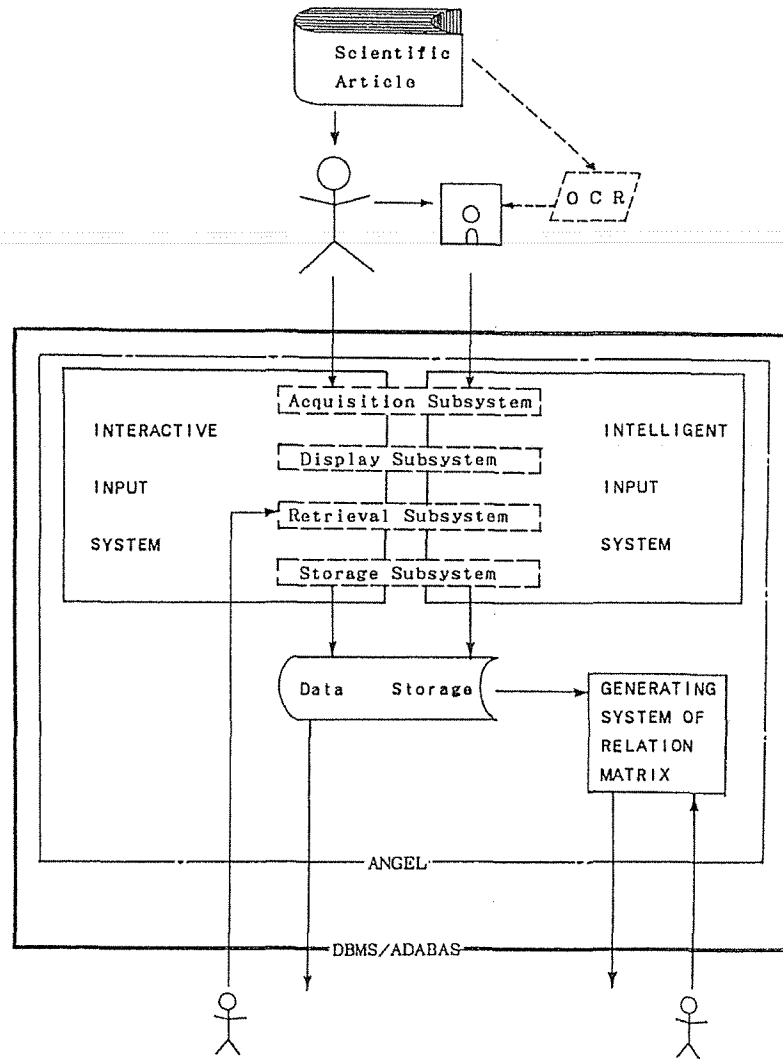


Fig. 1 The System constitution of the data base ANGEL.

specified relation and to output it into file.

### 3. Scientific Article Information Input System of the Data Base

#### 3.1 Acquisition of scientific article information

There is a choice to purchase a marketing article data base to acquire bibliographic items by extracting a data field mechanically according to the specified format, and such choice is popular in general. There is also a choice to input bibliographic items manually after printed strings of article is read automatically by an optical character reader. A present situation of data input into data base needs manual inputting until when the possibility of the automatic item extraction realizes and the latter choice is available practically.

The researcher data base needs as much as possible detailed information. Accordingly, sufficient information cannot be obtained only from the marketing bibliographic data base like the in former method, and then a system to input necessary information directly from an

Fig. 2 An example of the logging data in ANGEL.

```

MORE
PAGE 245
                                88-09-28 16:12:57
-----
FNAME  PROCESS  PASSWD  ARTNO  ARTNOLST  LASTUID  YYMMDD  TIMET
-----
CAD     STORE     3012    3012  X10044   88-08-23 11:53:54.6
CAD     STORE     3013    3013  X10044   88-08-23 11:54:35.7
CAD     UPDATE    -3014   3013  X10044   88-08-23 12:03:43.6
CAD     UPDATE    3013    3013  X10044   88-08-23 12:05:02.8
CAD     STORE     -3014   3014  X10044   88-08-23 12:06:30.5
CAD     DELETE    -3014   3014  X10044   88-08-23 12:08:23.7
CAD     UPDATE    3014    3013  X10044   88-08-23 12:12:58.0
CAD     UPDATE    3014    3013  X10044   88-08-23 12:14:58.2
CAD     STORE     3014    3014  X10044   88-08-23 12:15:20.5
CAD     STORE     3015    3015  X10044   88-08-23 12:15:36.7
CAD     STORE     3016    3016  X10044   88-08-23 12:16:03.3
CAD     STORE     3017    3017  X10044   88-08-23 12:16:20.6
CAD     STORE     3018    3018  X10044   88-08-23 12:16:44.4
CAD     STORE     3019    3019  X10044   88-08-23 12:17:35.6
CAD     STORE     3020    3020  X10044   88-08-23 12:17:54.3
CAD     STORE     3021    3021  X10044   88-08-23 12:18:14.2
CAD     UPDATE    3022    3021  X10044   88-08-23 12:48:35.8
CAD     STORE     3022    3022  X10044   88-08-23 12:48:50.2

```

original article is indispensable. A researcher cannot afford to the mounting many cost easily for inputting. Therefore, a researcher is required to input the data by himself, or the data is inputted by a part-time assistant. The system that enables to input data efficiently even if by nonprofessional key operator is necessary. For this reason, it is necessary to decrease the hitting frequency as well as possible at the time of data input in order to make key operation easy. Consequently, two systems using registered information in the data base were programed. The first system is interactive input and retrieval system, and the second is intelligent input system.

### 3. 2 Interactive input and retrieval system

When the input systems were designed, the following points were considered ;

1) Improvement of input working efficiency.

Especially, for a user who is not an professional key operator, to improve the efficiency of input work is effective.

2) Reduction of misinput rate.

In case of inputting abundant data, it is said generally that even professional key operators misinput at about 3 % rate. In consequence, the reduction of misinput is an important theme to acquire the information in order to minimize errors. So data were processed on the basis of the next point.

3) Valid use of registered data.

To consider an input processing or a retrieval processing for the researcher database of scientific article, both a parent (citing) article and a child (cited) article are equally significant, then it is necessary to treat them as independent articles because the database is to process various relations among articles. Accordingly, the same article appears many times as the reference which has a similar theme of research. Hence a tendency that the article by the same author appears frequently is remarkable in accordance with the increase of registered articles.

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

## LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

## FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

## E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.