

기계·재료(자본재)분야 검색엔진 개발, 운영 Development of search engine in machinery & materials fields



전 창 배

(KIMM 기계·재료기술정보센터)

'77 - '83 경북대학교 문헌정보학과(학사)
'83 - 현재 한국기계연구원 선임기술원



조 총 래

(KIMM 기계·재료기술정보센터)

'90 - '94 부산대학교 전자계산학과(학사)
'94 - '96 부산대학교 대학원 전자계산학과(석사)
'96 - 현재 한국기계연구원 연구원



김 용 호

(KIMM 기계·재료기술정보센터)

'89 - '93 부산대학교 전자계산학과(학사)
'93 - '95 부산대학교 대학원 전자계산학과(석사)
'95 - 현재 한국기계연구원 연구원

1. 개요

컴퓨터를 이용하여 정보를 관리하게 되면서 사용자가 원하는 정보를 정확하고 빠르게 찾기 위한 노력이 이루어져 왔다. 수많은 정보가 쌓이게 되면서 보다 효율적으로 정보를 관리하고 찾기 위해 나온 것이 정보검색 시스템이다.

1991년 웹의 등장으로 정보의 홍수 시대를 맞게 되었고, 웹 상의 방대한 량의 데이터에 대한 접근을 용이하게 하기 위해 정보검색 시스템을 적용한 것이 웹 검색엔진이다. 웹용 검색엔진은 로봇을 이용하여 웹 상에 흩어져 있는 수많은 정보들을 수집하여 데이터베이스화한다. 웹 검색엔진을 이용해서 자신이 찾고자 하는 웹 상의 데이터를 찾을 수 있게 되었다. 하지만, 아주 방대한 량의 정보들 중에서 사용자가 정말로 원하는 데이터만을 추출해 내기가 쉽지 않다. 현재 쓰이고 있는 모든 검색엔진들은 웹상에 구축된 정보의 실제 내용과 가치에 상관없이 모든 종류의 웹 데이터를 관리한다. 이는 전문적인 기술정보를 찾으려는 사용자에게는 아주 부적합한 시스템이다.

이러한 일반적인 검색 시스템의 단점을 보완하기 위해 기계·재료를 중심으로 한 자본재 기술정보와 관련된 데이터만을 전문적으로 수집하고, 이를 필요로 하는 사용자가 용이하게 전문기술정보에 접근할 수 있는 자본재 기술정보 전문검색엔진의 개발 필요성이 크게 대두되게 되었다.

현재 자본재 관련 기술정보를 찾기 위해서 이

용자들은 정보의 종류와 형태에 따라서 서로 다른 시스템에 접근하여 찾을 수밖에 없다. 규격, 특허 정보, 제품 정보, 개발 현황 정보 등 정보의 종류에 따라서 사용해야 하는 시스템들이 다르며, 정보 제공의 형태가 데이터베이스나, 웹을 이용한 html 문서이냐에 따라 찾는 방법이 제각각이다. 또한 다양한 형태의 정보검색 방법을 사용자가 모두 숙지하지도 못하므로 원하는 정보를 올바르게 찾는다는 것이 아주 어렵다.

또한 대부분의 검색 시스템은 여러 종류의 정보들을 모두 관리하는 범용 시스템이다. 이는 일반 정보들과 전문기술정보를 같이 관리하기 때문에 전문분야 정보검색을 하기가 어렵다. 즉, 전문 기술 용어와 일반 용어를 구분하지 않고 있고, 새롭게 나오는 전문 기술 용어에 대한 고려가 없기 때문에 정확한 정보 검색을 제공하지 못한다.

ICOMM에서 개발한 자본재 전문검색엔진은 이러한 어려움을 해결하기 위하여 자본재 관련 기술정보만을 전문적으로 구축한다. 또한 다양한 종류와 형태의 기술정보들을 통합 구축함으로써 사용자가 원하는 주제와 관련된 모든 형태의 기술정보를 동시에 제공할 것이다. 관련 산업 종사자들이 이용자들은 자신이 원하는 정보가 어디에 있을 지에 시간과 비용을 투자할 필요가 없으며, 한 주제에 대한 모든 형태의 정보를 한꺼번에 찾을 수 있다.

2. 개발 배경

정보화, 지방화시대에 지방과 수도권과의 편차 없는 정보전달체계를 확립하고 지방소재 기업체의 기술혁신 지원을 목적으로 '94년부터 수행한 지역정보화 사업의 성공결과로 4백여만건의 전문기술정보를 구축하게 되었다. 이를 지역 7백여 기업에서 지역통합망(ICOMM VAN)을 통하여 현장의 애로기술해결에 유용하게 활용하고 있어 서비스가 활성화되고 있다.

자본재 전문기술정보서비스는 경남지역 특화산업인 기계·재료분야 제조업체를 지원하기 위하여 구축된 국내 최대의 전문정보자원이 타지역 기업체로부터 월 20여만건 이상의 활용실적을 나타내게 됨에 따라 '98년부터는 본격적으로 전국적인 서비스에 돌입하게 되었으며, 검색엔진이 웹상의 전문검색도구로서 필요성이 절실하였기 때문에 개발에 착수하여 '98년 9월 1차 버전을 현재 서비스중에 있다.(<http://icommm.kimm.re.kr>)

3. 정보수집

전문분야의 정보서비스를 하기 위해서는 먼저 수집대상과 범위가 설정되고 이용자의 정보검색시 만족도를 높이기 위한 전략에 의한 정보수집이 뒤따라야 한다. ICOMM 검색엔진의 제공정보는 '94년부터 구축해 온 기존 기업정보DB, 제조상품DB, 연구결과DB, 규격DB등 16종의 데이터베이스로 구축한 정보와 관련 웹사이트의 정보로 크게 대별하여 하나의 검색엔진으로 동시에 검색토록 하였다.

3.1 기존 DB 정보 구축

'94년부터 정부의 시범사업으로 창원·마산을 중심으로 한 경남지역 특화산업인 기계·재료분야 산업체를 지원하기 위하여 약 4백만건의 기술정보를 DB로 구축하였다.

최초 text version으로 각 DB별 검색서비스를 제공하던 것을 '96년부터 Web version으로 16종의 DB를 통합검색형태로 서비스를 전환하였다. 이를 다시 검색엔진 개발을 통하여 웹사이트 수집정보와 같이 검색서비스를 하게 되었다.

3.2 기존 DB와 Web의 연동

웹 검색엔진에 의해서 데이터베이스를 검색하기 위해서는 웹과 데이터베이스를 연동시켜주는

데이터베이스 통로가 필요하다. 데이터베이스 통로는 데이터베이스 응용 프로그램이 WWW와 어떻게 연동되는가에 따라 여러 가지 형태가 있는데 이중 ICOMM 검색시스템에서는 CGI용 실행파일 방식을 사용하였다. CGI용 실행파일 방식의 구조는 그림 1과 같다.

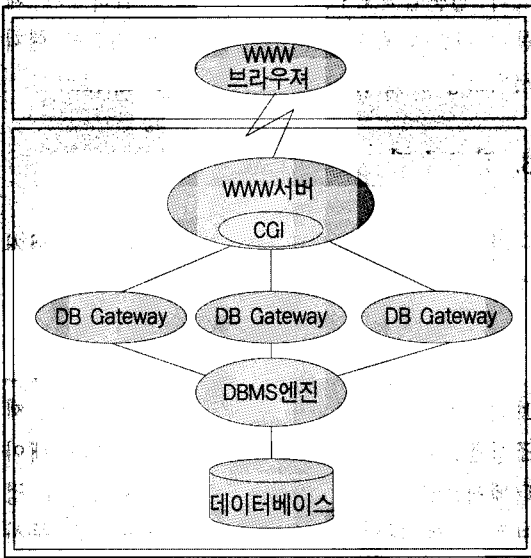


그림 1. CGI용 실행파일방식의 데이터베이스 통로 구조

CGI용 실행파일 방식은 크게 사용자들이 질의를 하고 결과를 보는 Client 부분과 사용자의 질의를 받아서 그에 대한 결과를 DB에서 찾아주는 Server 부분으로 나뉜다. Client쪽에서 사용자들은 WWW 브라우저를 사용하여 질의를 보내고 그 결과를 볼 수 있다. Server쪽에서는 WWW 서버가 CGI를 통해서 사용자의 질의를 데이터베이스 통로로 전달하면, 데이터베이스 통로는 사용자의 질의를 적당한 형태로 변환하여 데이터베이스 엔진으로 넘겨준다. 그러면 데이터베이스 엔진에서 수행된 질의의 결과가 다시 데이터베이스 통로를 통하여 WWW 서버에게 전달되고, WWW 서버는 그 결과를 HTML 파일 형식으로 Client쪽으로 보내주게 된다.

ICOMM 검색엔진에서 데이터베이스 통로의 역할은 데이터베이스 Key가 되는 제어번호를 입력

받아서 그에 대한 상세 데이터를 보여주는 것이다. 그림 2는 ICOMM검색시스템에서 데이터베이스 통로의 역할을 나타낸다. 이와 같은 데이터베이스 통로를 통해서 Robot Agency를 이용하여 ICOMM의 DB를 검색 엔진에 등록할 수 있다.

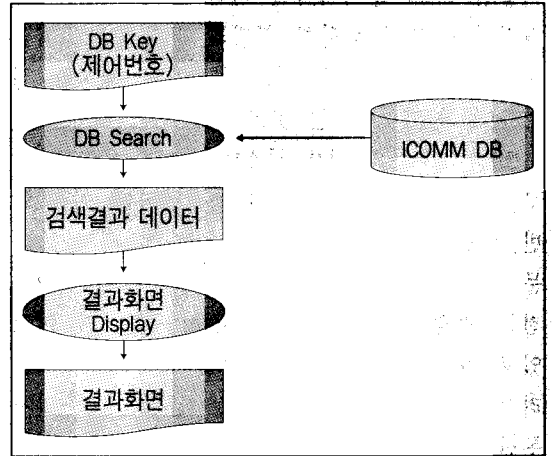


그림 2. 데이터베이스 통로의 역할

3.3 Web Site 정보선정 및 수집

일반 인터넷 검색엔진들은 웹상의 모든 사이트를 대상으로 Robot Agency를 방문시켜 모든 자료를 수집하게 된다. 그러나 전문기술분야의 정보만을 제공하기 위해서는 미리 방문사이트를 Robot에 지정을 해 주어야 한다.

ICOMM에서는 대상 사이트를 선정하기 위하여 Alta vista, Yahoo 등 기존 유수 인터넷 검색엔진을 이용하였다. 선정과정에서 가장 중요한 것은 과연 선정된 사이트에서 제공하는 정보가 산업체 기술개발에 얼마나 유용한 것인가와 관련 전문정보를 포함시켜 주는가 하는 망라성이 문제였다.

이를 위하여 국내 자본재 산업체들이 가장 필요로 기술정보가 무엇인가를 잘 반영해 줄 수 있는 기술개발 분야를 먼저 파악하는 것이 문제 해결의 관건이 되었다.

이를 현실적으로 가장 잘 대표할 수 있는 기준을 우선순위로 설정하였다. 첫째가 자본재분야

기업의 기술개발 수요를 반영하여 매년 발표하는 산업자원부 고시인 “자본재산업 전략품목 및 사업화품목”을 채택하였다. '98년 본 고시에는 핵심품목 62개, 일반품목 355개, 사업화품목 166개가 발표되어 있다.

각 품목에서 Keyword를 추출하여 우수 검색엔진을 통하여 해당 정보를 제공하는 사이트에

대한 List-up을 했다.

두 번째로 현재 국내에서 수행되고 있는 자본재분야 R&D 과제를 대상으로 하여 같은 방법으로 대상 사이트를 선정하였다.

두가지 기준에 의해 list-up된 사이트들을 다시 취합하고 분류분석을 통하여 중점수집 전략을 수립하였다.

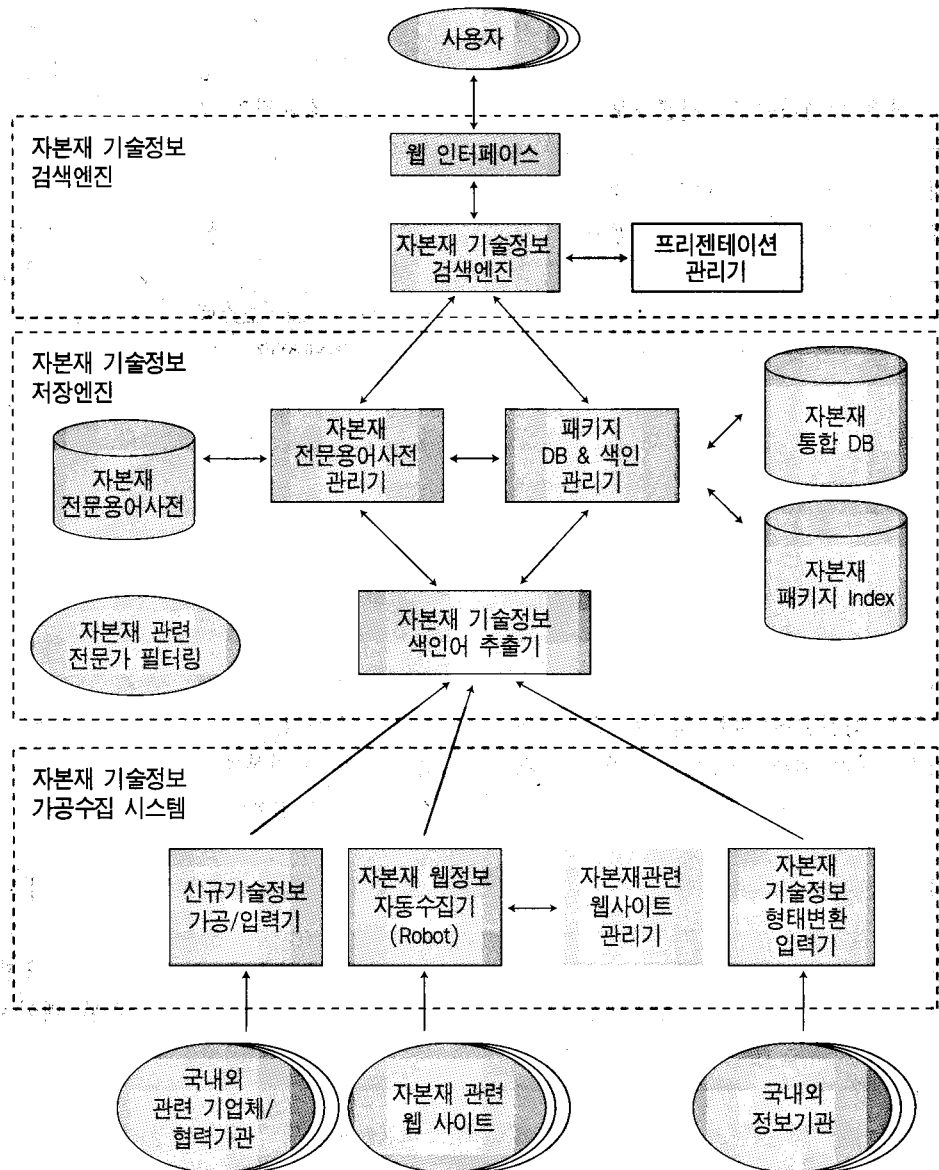


그림 3. 자본재 전문검색엔진시스템 전체 구성도

이들 사이트중 연구소(.re, .org등)와 회사(.com) 등은 해당 사이트 전체를 수집범위로 하였으며, 대학(.edu)등은 해당정보를 제공하는 하부 디렉토리 이하를 수집대상 사이트(Url)로 결정하여 Robot Agency 대상사이트에 포함시켰다.

4. 시스템 개발

본 전문검색엔진을 개발하기 위하여 다음의 4가지 분야로 서브시스템으로 구분하여 추진하였다.

현재 1차 버전 단계에는 1) 전문정보수집시스템, 2)저장엔진 시스템, 3)검색시스템을 개발, 적용하였으며, 4)전문용어사전 구축은 여러 가지 어려움으로 추후 적용키로 하였다. 전체시스템 구성도는 그림 3과 같다.

4.1 정보 수집, 가공시스템

자본재 기술정보 가공수집 시스템은 다양한 형태의 기술정보 소스를 수집하기 위한 서브시스템이다. 세 개의 정보 수집/입력기와 한 개의 웹사이트 관리기로 구성된다.

4.1.1 신규기술정보 가공/입력기

국내외의 자본재 관련 기업체와 협력 기관으로부터 신제품, 신기술에 대한 정보를 수집하여 데이터베이스화하기 위해 가공하고 입력하는 시스템이다.

4.1.2 자본재 웹정보 자동수집기

인터넷 상에 산재해 있는 방대한 량의 자본재 정보를 자동으로 수집하는 시스템이다.

4.1.3 자본재 웹사이트 관리기

자본재 관련 웹사이트를 발굴하여 집중관리하기 위한 시스템이다. 등록된 자본재 관련 웹사이트의 정보를 웹정보 자동수집기가 주기적으로 가져올 수 있게 한다.

4.1.4 자본재 기술정보 형태 변환/입력기

국내외의 타 정보기관에 의해 만들어진 자본재 관련 정보를 데이터베이스화하기 위해 정보의 형태를 변환하는 시스템이다.

4.2 자본재 기술정보 저장엔진

자본재 기술정보 저장엔진은 다양한 소스로부터의 자본재 기술정보를 자본재 전문용어사전과 전문가의 필터링을 통하여 통합 데이터베이스로 구축하고 효율적인 검색을 위한 색인을 만드는 서브시스템이다.

4.2.1 자본재 기술정보 색인어 추출기

각종 자본재 기술정보 소스를 데이터베이스에 입력할 때 그 소스에 대한 색인어를 추출하기 위한 시스템이다. 전문적이고 정확한 검색을 위해서 자본재 전문용어사전을 사용하여 다양한 방식의 색인을 지원해야 한다.

4.2.2 패키지 DB & 색인 관리기

자본재 기술정보 가공수집 서브시스템에 의해 수집된 정보를 관리하여 그에 대한 색인을 관리하는 시스템이다. 통합 데이터베이스와 색인은 관련된 자본재 기술정보를 한번에 검색할 수 있도록 구성된다. 또한 자본재 전문용어 사전을 이용한 색인은 자본재 관련 용어에 대해 정확하고 정밀한 정보 검색을 가능하게 한다.

4.3 자본재 기술정보 검색시스템

자본재 기술정보 검색엔진은 사용자의 검색요구를 정확히 분석하여 관련된 패키지 정보검색 및 일반 검색을 제공해 주는 서브시스템이다.

4.3.1 자본재 기술정보 전문 검색엔진

사용자가 입력한 질의어에 대해 자본재 전문용어 사전과 인덱스를 참조하여 관련된 모든 분

야의 정보를 한번에 패키지 형태로 제공한다. 정보 검색을 할 때 사용자는 특정 분야에 대해 한정하여 질의할 수도 있으며, 전체 분야에 대해 통합 검색을 할 수도 있다.

4.3.2 프리젠테이션 관리기

사용자의 요구 사항에 적합하도록 검색 결과 출력을 제어하는 관리기이다. 사용자의 요구에 따라 '일반검색엔진 결과 방식' 또는 '패키지 테이블 결과 방식' 또는 '게시판 방식'으로 결과를 보여준다.

4.3.3 웹 인터페이스

웹 환경에서 사용자가 검색을 할 수 있도록 웹서버와 검색엔진의 게이트웨이 역할을 수행한다. 사용자의 질의를 검색엔진에 전달하며 프리젠테이션 관리기에 의해 생성된 동적 html 결과를 웹서버로 전달한다.

4.4 자본재 전문 용어사전 관리 시스템

자본재 분야의 전문용어 사전을 관리하는 시스템이다.

4.4.1 자본재 전문용어사전 관리기

자본재 분야의 전문용어 사전을 관리하는 시스템으로서 신규 용어 입력, 용어 가중치 관리, 자본재 전문용어에 대해 동의어, 유사어 제공 기능을 한다. 자본재 기술정보 입력, 데이터베이스 구축시 인덱스 생성, 자본재 기술정보 패키지 검색 등 패키지 검색엔진의 다른 모든 관리 시스템에서 사용된다.

4.4.2 자본재 전문용어 사전

자본재와 관련된 기술용어를 사전으로 구축할 것이다. 자본재 관련 전문가의 필터링에 의해 구축한다.

5. ICOMM 검색엔진의 특징

ICOMM 검색엔진의 특징과 타 검색엔진과의 서비스 차별성은 다음과 같이 몇가지로 볼 수 있다.

첫째는, 기계·재료 및 자본재 분야에서 전문 기술정보만을 제공하되, 현시점으로 기술개발이 시급한 분야의 정보를 제공한다는 점이다. 이는 현재 국내 관련분야 R&D 수행과제와 450여 자본재 전략품목, 200여 무역역조 과다품목 및 중소기업 수요조사에 의한 대상기술을 Keyword하여 수집대상 정보 및 관련Site를 선정하여 수집, 제공하기 때문이다.

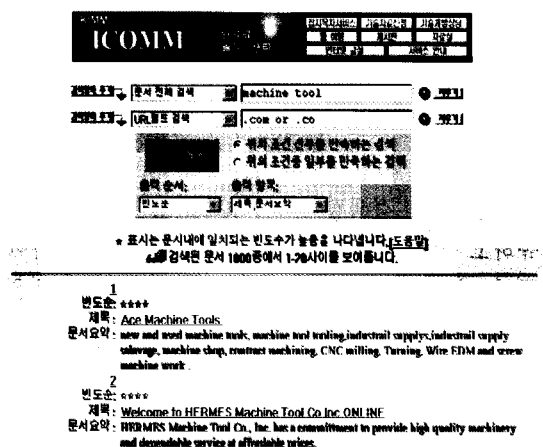


그림 4. ICOMM 검색화면

둘째는, 정보검색결과 제공초기에 이용자로 하여금 계속접근여부의 판단성을 제공한다. 이는 1차 검색결과 리스트에서 정보유형(기술논문·연구결과·기술보고서·규격·특허·기업/제품정보)정보를 제공하기 때문에 1차 검색결과만으로 계속 추적할 것인지를 대부분 결정할 수 있도록 해주며, 사전에 정보유형을 한정하여 검색할 수 있도록 되어 있다.

세째는, 검색시스템이 쉽고 유연성을 가지고 있다. 이는 검색자가 머릿속에 생각하고 있는 내용을 검색시스템에서 충분히 표현할 수 있게 함으로써 정확한 정보를 찾아내는데 큰 도움이 된

- 1
 빈도순: 100000
 제목: **Microbeam Synchrotron Radiation...ase Turbine Blade After Service**
 문서요약: Microbeam Synchrotron Radiation Diffraction Study of a Monocrystalline Nickel-Dase Turbine Blade After Service 김지복차서비스 원문 복사 신청 Microbeam Synchrotron Radiation Diffraction Study of a...
- 2
 빈도순: 100000
 제목: **[구격] Acceptance Criteria for ...ine Blade Grain Size (Proposed)**
 문서요약: [구격] Acceptance Criteria for Cast Turbine Blade Grain Size (Proposed) SAE_TYPE S DOC NUMBER SAE/J 2160 초록 Abstract: There is no Abstract for this document. Publication Year: 1990
- 3
 빈도순: 100000
 제목: **[기술논문] Extending turbine blade life**
 문서요약: [기술논문] Extending turbine blade life SAE_TYPE M DOC NUMBER 2-11-2-27C 초록 Authors and Their Organizations: Birch, Stuart Society of Automotive Engineers, Inc. Abstract: There is no Abstract for
- 4
 빈도순: 100000
 제목: **[기술보고서] The role of the tu...**
 문서요약: [기술보고서] The role of the turbulent prandt number in turbine blade heat transfer prediction SAE_TYPE P DOC NUMBER 901810 (SAE TRANS., VOL. 990) 초록 Authors and Their Organizations: Whitaker,...
- 5
 빈도순: 100000
 제목: **[기술보고서] Turbine blade heat...**
 문서요약: [기술보고서] Turbine blade heat transfer prediction using coupled turbulent Prandt number and transition models SAE_TYPE P DOC NUMBER 912152 초록 Authors and Their Organizations: Whitaker, Kevin W

그림 5. 검색결과

다. 별도의 연산자 조합없이도 검색항목 추가창을 통하여 Keyword만 잘 입력하면 원하는 결과를 얻을 수 있으며, 정확성이 높은 정보순으로 결과를 보여주기 때문에 만족도를 높여준다.

또한 최종 검색결과에서 ICOMM이 원문을 보유하고 있는 기술자료는 “원문신청”이라는 아이콘이 나타나며, 이를 이용하여 바로 원문주문을 할 수 있도록 되어 있다.

6. 발전계획

ICOMM의 자본재 검색엔진 국내 취약산업에 대한 기술개발과 생산제품의 국제경쟁력 향상에

기구불바를 이용한 공작기계의 오차평가 시스템 개발
 Development of Computer Aided System for Error Assessment for Multi-axis Machine Tools using the Double Ball Bar

원문신청

| | |
|--------|------------------|
| 국내외구분 | 국내 |
| 자료구분 | J |
| 언어코드 | KOR |
| 소장기관명 | 한국기계연구원 |
| 지차명 | 전통회 |
| 저자소속기관 | 서울대학교 대학원 기계설계학과 |
| 출판사 | 한국정밀공학회 |
| 출판지 | 서울 |
| 출판년월 | 1998 |

그림 6. 원문주문 화면

일익을 담당할 수 있을 것으로 판단하고 있다. 앞으로 보다 많은 전문정보자원의 확충과 고품질 서비스를 위한 많은 노력이 뒤따라야 할 것으로 본다.

본 논문은 국내 연구소, 대학 등에서도 앞으로 뚜렷한 목적성을 가진 정보서비스가 이루어져야 할 것으로 판단하고 있기 때문에 금번에 ICOMM에서 시도한 전문분야 검색엔진 서비스가 많은 도움이 되었으면 하는 바램이다.

참 고 문 헌

- [1] 통상산업부 고시 제 1998-21호, 1998. 2. 27
- [2] 전형덕 외 “창원·마산 지역정보유통망 구축” 한국기계연구원 기계재료기술정보 센터 연구보고서, 1996. 12.
- [3] 전창배 “한국기계연구원 기계·재료기술 정보센터” 정보관리연구 28(1), pp.90-106, 1997. 3.