

新聞資料 데이터베이스에 관한 고찰

Database for Newspaper

남상식*

초 록

신문의 자율화시대가 왔다. 정부로부터의 규제가 풀림으로 해서 신문사는 자유경쟁속에서 흥하던가 아니면 망하게 되는 적자생존의 시대가 90년대 초반에 닥칠것으로 예견된다. 산업으로서의 신문업은 신문사경비의 대부분을 차지하고 있는 인건비라는 약체성 때문에 본업을 보완할 필요성이 절실하다. 따라서 신문제작의 전산화는 인건비 절약면에서 크게 기여할 수 있으리라고 본다.

ABSTRACT

Computerization of Newspaper production will help reducing a tremendous labor cost that has been a serious element for survival of Newspaper industry.

The purpose of this paper is to emphasize a urgent need for the construction of Newspaper database.

1. 서 언

신문의 質을 평가하는 기준은 “정확성”이다. 정확성은 오류가 없는 축적 정보에서 나오며 사회현상중 인간의 지각과 판단에 의해 선택된 지식이다. 신문이 비교적 가공된 지식을 전달한다는 논리는 이렇게 시작된다.

신문 자율화 시대가 되었다. 정부로서는 규제

를 다 풀었으니 자유경쟁속에서 신문사끼리 이 전투구해 몇몇은 흥하고 또다른 몇은 망하게 되는 적자생존의 시대가 90년대 초반에 닥칠 것이라 예견하고 있다. 산업으로서 신문업이 본업의 약체성 때문에 본업을 보완할 필요가 생기게 될 것이라는 논리도 된다. 왜 본업이 약체가 될까. 그 이유는 간단하다. 인건비가 신문사 경비의 많은 부분을 점유하기 때문이다.

* 동아일보 조사부

신문사의 기본 업무는 정보의 수집에서부터 편집, 인쇄, 발송을 통해 각 가정이나 직장에 배달되는 시스템이다. 신문업의 처음과 끝이라 할 수 있는 정보수집과 배달은 기계화 할 수 없으며 따라서 사람의 손에 의지할 수 밖에 없다. 이 때문에 신문업은 인건비가 원천적으로 많이 든다. 한마디로 신문업은 지적 노동력에 의존하는 노동 집약산업으로 인건비 비율이 전체 코스트의 1/3 이상을 점유한다.

인건비 수준도 굉장히 높다. 동아일보나 조선일보 같은 국내 최대의 신문사의 경우 수습을 마친 기자의 초봉이 65만원선이며 업무직도 비슷한 수준으로 국내 유수의 대기업체 과장 봉급보다 많다. 과장이 되려면 5-6년 걸리는데 단 6개월만에 그만큼 봉급을 받게 된다. 더우기 장기 근속자가 많아 인건비의 지출은 앞으로 계속 증가할 전망이다. 게다가 언론 노조의 활성화로 인해 임금 수준은 계속적으로 상향을 거듭할 것이 분명하다.

갈수록 체질이 약해질 신문사가 치열한 경쟁 속에서 살아남기 위해서는 인건비를 줄이고 본업 외에 부업을 늘려갈 수 밖에 없다. 그러기 위해서는 신문의 제작기술 혁신이 필요할 것이고 이를 위한 투자도 장기적으로 계속 되어야 할 것이다. 문선공이 활자를 한개씩 뽑아 식자공이 조판하고 지형을 만들어 연판을 구워내 운전기에 걸어 인쇄하려면 많은 사람, 많은 인건비가 들게 된다. 이 공정을 전자동 모노타이프 -- 팩시밀리 -- CTS -- 전자편집의 순으로 기술을 발달시켜 사람을 줄여야 한다.

전자동 모노타이프와 팩시밀리는 1960년대 일본 신문사에서 성행했으며 CTS는 70년대에, 80년대 요즘에는 새로운 전자기술을 도입, 전자 편집 시스템을 상용해 인건비를 혁신적으로

떨어뜨리고 있다. 전 산업면에서 약 10여년이 뒤떨어진 우리 실정으로는 컴퓨터를 사용하여 VDT 상에서 직접 편집, 조판을 하는 CTS(Computerized Typesetting System) 방식이 가장 적합할 것으로 본다. 신문제작이 CTS화하면 社當 1천명선을 유지하는 국내 6대 일간지의 인원이 社當 약 1백 50명 줄어듦 전망이다.

금전적으로 계산해 본다면 신문사 문선, 식자, 조판, 인쇄공 1인 평균 임금은 연간 1천만원 수준이다. 1백 50명이면 적어도 연간 15억원의 인건비 절감과 대략 5억원의 비용절감 효과를 가져볼 수 있다. 한편, CTS에 소요되는 예산은 2-3백억원 가량 된다. 연간 15억원의 임금이 매년 10%씩 상승하고 비용이 그 이상 증가한다고 가정하면 단 5년만 지나도 1백억원 이상을 절약할 수 있으며 CTS와 함께 개발되는 데이터베이스를 활용, 정보산업이라는 부업까지 한다면 5년간 약 1백 30억 - 1백 50억원을 절감할 수 있게 된다. 이런 계산으로라면 동일한 조건의 신문사 중 CTS를 한 신문사와 하지 않은 신문사는 부가가치를 감안, 적어도 7년이면 승부가 나게 된다. CTS를 해 놓은 신문사는 7년 후 적은 인원으로 더 신속하게 신문을 제작할 수 있게 되고 그렇지 못한 신문사는 여전히 노동 집약적인 상태를 벗어나지 못해 경영이 부실해질 수 밖에 없는 것이다.

CTS는 신문 제작의 전산화다. 신문 제작 과정 중 신문기사를 쓰는 순간부터 전산화 된다. 그러므로 신문기사 전산화는 CTS와 함으로써 입력이 자연적으로 이루어질 수 있다. 즉 CTS와 신문기사 데이터베이스를 동시에 수행한다는 것이다. 그러나 반대의 경우도 있을 수 있다. CTS가 너무 예산이 많이 들고 고도의 기술이

필요하기 때문에 예산이 덜 들고 비교적 논리적 기술이 필요한 신문자료데이터베이스를 착수, 정보산업이란 부업부터 개시한 후 축적 기술을 바탕으로 CTS를 하는 경우도 그에 못지 않다. 어떤 경우든 컴퓨터는 인건비 절약면에서 위력을 발휘할 수 있기 때문에 신문자료의 전산화는 절실하다는 인식이 팽배해 있다.

2. 新聞資料의 本質

신문제작에 필요한 자료는 형태상 7가지로 구분된다. 기사 스크랩, 사진, 인물정보, 마이크로物, 定刊物색인, 도서, 定刊物이 바로 그것. 가장 접근이 용이하고 이용이 빈번한 것은 스크랩된 기사와 사진 그리고 人的사항이 망라된 인물정보다.

그리고 여기서 해결되지 못하거나 좀 더 자세한 정보를 원하는 이용자들에게 도서와 정기간행물을 지원하게 된다. 물론 정기간행물색인은 후자의 이용에 있어 기본적 TOOL이 된다. 자료형태가 어떻든 신문제작에 직접 혹은 간접적으로 지대한 영향력을 미치는 것은 사실이다.

일반적으로 각 자료들은 나름대로 지면에 반영하는 형상이 다르다. 스트레이트기사에는 기사스크랩, 사진, 주제 파일 등이 주로 이용되고 해설기사나 기획기사는 도서, 정기간행물이 주종을 이루고 있다. 즉 대부분 스트레이트性 여부에 따라 주요형태가 달라진다고 볼수 있다.

한편 단답형 질문에는 신속히 제공되지만 하면 되기 때문에 자료 유형에 구애를 받지 않지만 논리 해설형 질문에는 각종 자료를 교차시켜 제공한다. 논리해설형 질문은 대개 기획기사 혹은 특정 주제의 社說用 참고 자료를 얻고자 던져지는데 워낙 전문적이라 검색에 애로점이 많

은 편이다.

이상과 같이 신문자료는 제한된 시간내에 지원돼야 하기 때문에 신속성이 있어야하고 자료의 상치된 바(통계 또는 수치)를 올바르게 제공해야 하는 점 때문에 정확한 근거를 제시해야 하며 상반된 주장을 동시에 제공할 필요가 있어 다양한 논지를 펴 주어야 한다.

3. 利用者 分析

신문자료이용자는 취재기자가 대부분이다. 휴간 자녀의 숙제 해결이나 업무상 참고자료가 필요한 사람들이 이용하기는 하나 1% 미만이다. 이밖에 외부이용자들도 적잖이 많다. 즉 논문 쓰려는 학생들의 주제파일 이용 요구, 초중고 학생들의 과제물 해결 요구, 독립유공자 지정용 자료 요구, 특정 인물의 인적사항 및 소재확인 요구, 기사검색 요구 등 사회전체에서 골고루 자료이용을 강력히 요청하고 있다.

주 이용자인 취재기자는 크게 두가지로 이용형태가 구분된다. 하나는 사전취재, 즉 취재를 위한 사전 취재를 하는 경우고 다른 하나는 취재후 사후 보충 취재다. 사전취재는 철저히 자료에 의거한다. 취재대상에 한번도 접근하지 않은채 모든 정보를 거머쥘수 있고 때에따라 기사의 대부분을 마쳐놓고 취재하러 가는 경우도 왕왕 있다. 사후취재는 철저한 보완에 한정된다. 사후취재는 철저한 보완에 한정된다. 다시 말해 통계나 수치정보의 보완, 부정확한 사실의 확인 등이 대부분이다. 특수한 경우 취재원 접근이 되지 못해 취재자체가 이루어지지 못했을 경우도 사후취재를 하게된다.

최근 신문자료가 법적증거물, 최신통계인용, 생활정보취득, 생활경제정보 습득 등에도 효과

적으로 활용되면서 각 방면으로 이용자가 늘어 가는 추세인데 이에 대한 대책으로 사회의 公器로 불리는 신문사의 보다 적극적인 정보공급 태도가 주목되고 있는 실정이다.

4. 신문자료전산화의 기본적인 입장

신문에는 대체로 보도, 논평, 오락의 3 가지 기능이 있다. 이 중 오락면을 제외하고는 사회 현상에 관련된 사실과 이에 대한 의견개진, 해설등을 지면에 실는것이 신문의 주요 역할이다.

신문정보가 모두 “사실” 자체일 수는 없다. 그러나 신문 정보에 대신할 만한 정보를 대중에게 비교적 완벽히 전달할 단체는 기술적으로 존재하기 힘들다. 그래서 신문을 “재현된 세계상”으로 표현하기도 하고 “역사”를 대변하는 연구 소재로도 평가하고 있다. 신문정보는 이런 가치 때문에 연구 대상이 되고 있는 것이다.

신문 정보는 직접 취재와 자료 보충으로 완성된다. 자료는 사전취재와 취재후 어느 시기에나 활용되고 있다. 자료 축적이 기하급수적으로 되자 국내에서도 80년대 이후 사람 손으로 아무리 효율적인 분석을 해서 자료를 정리한다고 해도 사후 검색에 애로가 늘 있어 미처 손도 못대는 사이 어느새 미정리 자료가 또 쌓여 가는 현실을 해결하기 위한 시도가 이루어지고 있다.

일본의 경우 올림픽을 전후해서 신문들이 일제히 증편을 단행한 일이 있었는데 그 현상과 부작용이 최근 국내에 밀어닥친 것이다. 증편하면 사전 사후 취재에 자료가 많이 필요해져 자연 검색상의 애로를 해결하기 위해 컴퓨터를 찾게 된다. 자료의 전산화는 바로 이런 필요성에서 출발한다. 필자의 경험상 자료 전산의 절차는 다음의 10 단계로 설명할 수 있겠다.

- 1 단계 준비 위원회 구성
- 2 단계 대상 자료 결정, 마스터플랜 확정
- 3 단계 인원 및 예산 지원
- 4 단계 대상자료 조사 분석(모델 작성 포함)
- 5 단계 시스템 설계
- 6 단계 프로그램 작성

- 7 단계 테스트 런
- 8 단계 시스템 가동
- 9 단계 시스템 평가와 유지
- 10 단계 서비스 개시

1 단계는 여러모로 중요한 의미를 갖는다. 자료 전산이 도대체 어떻게 진행되어야 하는지, 꼭 해야 할 필요성이 있는지를 면밀히 검토하는 단계이기 때문이다.

준비위원회는 관련 간부진과 전문가들의 협의체로 구성하고 별도로 실무 연구회를 운영 하는 것이 관례로 되어 있다. 즉 준비 위원회에서는 큰 계획을 검토하고 그에 부속된 실무 연구회는 구체적인 사항을 세세히 고찰해야 한다. 준비 위원회가 검토할 사항은 다음과 같다.

- 1) 자료의 실태를 파악하는 실태 조사와 그에 따른 업무 분석
- 2) 어떤 자료의 전산화가 이루어져야 하는 필요성과 타당성 조사
- 3) 전산화 대상 자료의 경제성 여부
- 4) 타사와 외국의 자료전산 실태와 활용상황
- 5) 어떤 단계를 밟아 전산화를 시도할 것인가?
- 6) 대상 자료의 우선 순위는 어떻게 할까?
- 7) 입력 자료를 만들 인원과 이를 관장할 데이터 프로듀서는 있나?
- 8) 전산 실무진의 기술 축적은 만족할 만한가?
- 9) 예산 지원은 원만한가?
- 10) 자료전산을 실행할 공간은 어떻한가?

11) 전담조직은 어떻게 구성할 것인가?

12) 회사의 전반적인 계획과 어긋나는 점은 없나?

13) DB 구축은 어떻게 할 것이며 국내외 데이터뱅크와 연결은 가능한가?

14) 자료전산을 함으로써 얻어지는 회사의 이익은 무엇인가?

15) 전산화된 정보의 판매는 어떻게 할 것인가?

이상의 요소들은 면밀히 검토하여 전산화 방향을 명확히 제시하고 그에 따라 청사진을 제시하며 기존팀과 외부연구팀이 합동으로 마스터플랜을 작성해야 한다.

2 단계는 대상 자료의 결정이다. 신문 제작에 필요한 갖가지 자료 중 어떤 자료를 전산화 대상으로 할 것이며, 대상이 된 자료는 우선 순위를 정할 것인가 아니면 전체를 동시에 DB화할 것인가를 결정해야 한다.

마스터플랜에서 2 단계 과정이 결정되는 경우도 왕왕 있기는 하나 현업에서 실제로 부딪히는 실무진에게는 2 단계가 에러를 줄이기 위한 필수적인 과정이기도 하다.

3 단계에서는 정책적인 뒷받침이 거론되어야 한다. 마스터플랜에서 제시된 예산과 전산화 연차 계획에 추가로 소요될 예산을 이사회 혹은 이에 준하는 기구에서 인준해 주는 절차다. 전산화는 예산과 인력이 뒷받침되지 않는 한 추진하기가 매우 힘들기 때문이다.

일단 전산화 계획이 승인되고 예산안이 통과 되면 최소한 개발이 끝나기까지는 연구팀으로 하여금 소신껏 작업을 할 수 있도록 정책적인 배려도 해야 한다.

4 단계부터는 데이터베이스화와 함께 다음장에서 설명한다.

5. 新聞資料데이터베이스화

각종 자료중 전산화 대상을 선정한다는 것은 쉬운 일이 아니다. 고려해야할 요소들만 해도 이용빈도, 필요성, 경제성 등이 있고 기술적인 면에서 난이도를 정확히 평가, 해결가능한 가이드라인을 설정해야하는 번거로운 일이 기다리고 있기 때문이다.

전장에서 열거한 자료 형태중 난이도만을 기준해 난이도순으로 순위를 매기자면 프로필(인물정보), 정간물색인(신문 제외), 도서와 정간물, 사진, 마이크로物, 신문기사색인, 신문기사순이 될것 이다.

프로필은 검색대상 아이টে을 고정시키고 아이টে을간 논리검색을 하면 비교적 전산화가 용이해진다. 정간물색인에서는 대부분 통제어휘를 사용하기 때문에 검색대상 주제어 선별이 까다롭지는 않다. 다만 통제어를 자동색인케 하거나 유사어간 표준 통제어를 사전에 만들어 두면 한결 편해진다. 문제점이 있다면 한자처리다.

도서 정간물은 구입요청, 입수, 분류, 목록, 대출, 반납까지 TOTAL SYSTEM이 바람직하므로 신문사에서 직접 손대기 보다는 관계 전문가들에 일임하는 편이 좋을 듯하다. 마이크로物은 70년대부터 활용해온 COM시스템과 블립을 넣어 찍은 자동검색 시스템이 성행하나 현실적으로 수요가 적은 신문사 실정으로 보아 경제성이 없다고 판단된다. 최근 CD-ROM 기술을 이용해 마이크로物을 CD로 대체하는 경향이 일반화 되었으나 아직 우리 신문사 수준에서는 경제성 때문에 검토되기가 힘든 실정이다. 다만 수요가 늘게되면 CD로 대체할 연구와 기술습득을 게을리해서는 안될 것이다.

신문기사와 신문기사 색인은 주제어의 통제나

아니냐를 결정하는 단계서부터 문제가 야기될 수 있다. 이 부분은 추후 케이스 스타디를 첨부, 효율적인 방안을 제시할 예정이다.

신문기사는 작성 자체가 RANDOM하고 이용도 구 역시 마찬가지로 때문에 기억을 의지한 수작업으로 완벽한 논리 검색은 불가능하다. 특히 어느 한 WORDING 만을 발췌 요구할 때는 더욱 그렇다. 최근 초이용자들의 컴퓨터를 사용한 정보이용요구는 이러한 수작업의 한계를 직시하는 것으로 사료되는바 신문기사전산화가 시대적인 추세임을 부정할 수 없다.

신문기사의 전산화 과정은 크게 3 단계로 나누어 볼 수 있다. 1 단계는 설계 단계로 기사 유형에 따른 입력 Layout 이 실행되며, 2 단계는 입력단계로 S/W과 H/W를 매치시켜 정형화된 파일을 축적시키는 단계다. 3 단계는 검색 단계로 입력이 완료된 신문기사를 이용요구에 따라 논리적인 검색과 이용 관행을 축적시켜 가는 과정이다.

신문기사 건수가 1 백만을 넘게되면 아무리 유능한 정리자라 하더라도 수작업으로는 요구에 적합하도록 검색해 내기가 어렵다. 논리적으로 복합적인 기사 요청 역시 수작업으로는 거의 불가능하다. 비단 검색뿐 아니라 정보화 사회에서의 발달된 전자 기술이 인원을 줄이는데 있어 신문기사를 핸들링하는 부서라고 예외로 놔두지는 않는다. 컴퓨터를 사용함으로써 루틴 업무를 전산처리하고 남는 시간을 다른 업무를 볼 수 있게 하는 것이다. 그러나 신문기사를 전산화할 때 초기 작업인원은 많아질 수 밖에 없다. 그래서 전산화의 경제성을 고려, 가장 적은 비용으로 가장 높은 부가가치를 창출해야 한다.

그러기 위해 설계 단계부터 고려할 사항은 많아질 수 밖에 없다. 고려할 사항은 다음과 같다.

1) 신문기사의 출고 후 판별로 분화되는 각종 기사의 기준은 무엇으로 할 것인가

2) 지면 전체를 입력할 것인가, 아니면 지면 일부는 삭제할 것인가

3) 수록기간은 어느 정도로 할 것인가

4) 신문 기사만 할 것인가, 아니면 광고, 안내 등 지면 일부는 삭제할 것인가

5) 회사의 전반적인 전산 계획과 어긋나는 점은 없는가

6) 소프트웨어 개발은 어떻게 할 것인가

7) 他 데이터베이스와의 연결은 어떻게 할 것인가

8) 상용 데이터베이스로 할 것인가, 자체 이용 데이터베이스로 할 것인가

9) 디소러스를 쓸 것인가 Free Term 방식을 채택할 것인가

10) 디소러스를 쓴다면 디소러스 준비팀의 노우하우나 제작기술이 있는가

11) 시스템은 무엇으로 할 것인가

12) 파일럿 시스템을 도입할 것인가 Test-run을 당분간 유지할 것인가 이상의 요소들을 면밀히 검토, 구축방향을 명확히 제시하고 그에 따라 청사진을 제시해 외부 연구팀이 합동으로 마스터 플랜을 작성해야 1 단계인 입력 준비가 끝나게 된다.

2 단계인 입력단계 전에 또 할일이 있다. 혹 설계의 잘못으로 시스템 전체를 바꿔야하는 불상사가 생길 수도 있기 때문이다.

하나의 시스템이 구성되는 기본적인 요소는 입력(Input), 출력(Output), 처리(Process), 피드백, 제어(Control)다. 설계의 순서는 가) 대상자료나 정보의 파악 나) 출력설계와 입력설계 다) 처리 과정 설계의 3 절차를 밝게 된다.

첫째 절차는 초기의 조사 분석을 구체화 하는 작업이며 둘째 절차에 들어가기 전에 원자료(Source Data)를 완벽히 준비해야 한다. 만약 원자료의 구성이 복잡하면 이용에 편리하도록 데이터를 가공하기도 한다.

둘째 절차에서는 이용 형태에 따른 입력 포맷을 레이아웃하고 아울러 출력 형태로 지정해야 한다. 보통 말하는 워크시트(Work Sheet)는 입력 포맷에 해당되며 검색 대상이 되는 아이템과 그 아이텍끼리의 논리 검색을 겨냥, 파일을 고정 필드와 가변장 필드로 분리해야 한다.

세째 절차는 설정된 자료나 정보를 컴퓨터로 처리하는 과정, 즉 정보간의 상관 관계와 정보의 사용 범위를 확실히 하기 위해 자료의 결합 관계를 파악하는 일이다. 이 세 절차를 끝내면 자연히 시스템 플로우차트를 작성할 수 있으며 전산 작업의 전체를 관장할 수 있게 된다.

세 절차를 수행하는 동안 보다 중요시되는 점은 시스템 관리 측면보다 검색 측면이다. 정보의 저장과 유지에 치중하는 시스템관리보다는 검색 즉 원하는 정보를 찾는 것이 급선부. 따라서 정보 검색 시스템을 보다 정밀히 하는 것이 관건이 될 것이다. 시스템 설계가 완료되려면 하드웨어가 결정돼야 한다. 하드웨어는 소프트웨어와 데이터량을 충분히 커버할 만큼의 용량이 필요하다. 차후 늘어날 데이터 수량도 감안해서 결정해야 한다. 아울러 응답 시간의 신속성도 심각히 고려해야 한다. 최근 추세는 시스템에 소장된 자료가 최대 용량이 아닌 이상 하드웨어를 구입하지 않고 임차해서 쓰는 방법도 간혹 나오고 있다.

그 다음은 프로그램 작성이다. 프로그램은 기존 프로그램을 쓰는 방법과 자체 개발한 프로그램을 사용하는 방법이 있다. 자료 전산의 경험

이 전혀 없는 상태가 아닌 언론사 경우 두 가지 방법을 혼용하는 것이 효과적이긴 하지만 기존 프로그램들이 코드문제등 아직까지는 제약이 많아 도입에 따른 문제점이 돌출될 우려도 없지 않다. 프로그램이 작성되면 하드웨어와 소프트웨어를 매치시키고 전단계에서 설계된 시스템을 거기에 실어 보는 테스트 런이 실행된다. 테스트런 과정중 발생하는 에러를 교정하고 입력 자료를 재편집하는 등의 마무리 과정까지 함께 처리하게 된다.

신문기사의 전산화는 단순히 전산화에만 그치지 않고 종류가 다른 자료와도 상호 논리검색이 가능하도록 데이터베이스화하는 편이 좋다. 사실 몇년 전까지만 해도 신문기사 DB는 일반화된 얘기가 아니었다. 그러나 일본신문사들이 60년대말부터 앞을 다투어 개발을 시작, 80년대에 이르자 대부분의 신문사들이 개발을 완료, 본격적인 서비스와 상용화를 개시하자 국내에도 그 영향이 파급되었던 것.

신문기사의 DB화는 3 가지 방법을 제시할 수 있다. 우선 형태별로는 일부주제 분야의와 전체분야, 시차별로는 소급방식과 사후방식, 형식별로는 발췌형식과 전문형식이 있다. 이상적인 DB라면 전체분야를 소급하여 전문과 초록을 동시에 나타내주는 것인데 일본이나 구미의 경우 대부분 이방식을 채택하고 있다.

DB방향이 결정되었다면 다음 작업은 SOURCE 데이터인 신문기사의 입력형태를 경제성을 고려, 고정필드와 가변필드로 나누어 입력하는 것이다.

고정필드에는 게재일, 신문명, 面, 版, 段, 필자명, 제목, 부제목, 분류어, 키워드 등의 검색 KEY가 반드시 배정되어야 하고 가변필드는 초록이나 전문이 들어가게 된다. 초록은 2 백자 내지

4백자의 기사요약이 일반적이며 뉴스플래시 경우는 50 자부터 1백자까지가 대부분.

신문기사의 DB화는 짧은 제작시간내에 다량의 정보를 다수에게 제공해야 하고 동시다수에게 제공되어야 한다는 점에서 필수적이다. 이런 점에서 보면 종래의 파일을 계속 사용해 온 단말 사용자들이 즉석에서 원하는 정보를 제공받고 쉽게 파일취급 방법을 익힐수 있게된다. 그러나 DB 파일 구성과 범용프로그래밍 언어를 정확히 알지 못하는 단말용자들을 위해 질의의(QL) 교육등 기본교육은 있어야 한다.

시스템 문제는 지난 10년동안 가장 관심있게 연구돼 온 데이터베이스 컴퓨터(DBC)의 채택도 검토할 만하다. DBC는 호스트컴퓨터, 마이크로컴퓨터까지 이용할 수 있으며 필요에 따라 이들의 혼합형도 설치할 수 있다. DBC는 또 데이터베이스 저장장치에서 원하는 데이터를 찾아 호스트컴퓨터에게 넘겨 주기때문에 호스트컴퓨터는 사용자와 DBC간의 인터페이스 역할도 해주게 된다.

한편 데이터베이스의 형태문제도 요즘 심심찮게 거론된다. DACOM의 DNS가 분산 데이터베이스의 활용을 시작한 후 보급된 온라인분산 데이터베이스는 지리적인 데이터의 저장, 통신 시스템이 가미된 네트워크를 이용, 그 구축진도가 상당히 진보적이다. 신문기사가 아닌 인물정보 수준, 혹은 목록등 도서관리 업무수준은 PC로 LAN을 구성, 처리하는 방안도 검토될 수 있으나 용량 수급문제는 여전히 미지수다.

6. 국내의 사례

일본 신문사의 신문기사 DB는 일본경제신문이 효시다. 日經은 1968년 부터 CTS와 더불어

어 데이터베이스를 동시 개발에 착수했고 朝日, 讀賣 등은 1970년 후반에, 비슷한 시기에 中日 静岡新聞등의 지방지가 개발에 착수했다.

* 일본경제신문

1975년 5월부터 데이터 축적을 시작, 국내외의 매크로인 정치, 경제 동향으로 부터 각 산업, 기업의 동향에 이르기까지 폭넓게 커버한다. 정보의 특징은 닷케이 신문의 지방 경제면을 수록하고 있고 닷케이 이외의 신문에서는 화학공업일보를 추가했다.

DB명은 NIKKEI로 수록 건수는 약 70 만건 기사정보의 형태는 40 자 이내의 기사 요약 및 전문, 기사 초록, 지면 마이크로피쉬 등 다양하다.

* 中日新聞

1972년 1월 축쇄판의 기사색인지인 월간 中日新聞 색인집 발간후 74년 12월 뉴스 이소러스 초판, 78년 10월 데이터뱅크 인덱스, 81년 1월 뉴스 디소러스 중보 개정판을 내후 82년 4월 지방판 뉴스 데이터 뱅크를 구축했다. 中日는 72년 부터 81년까지 10여년간 DB연구에 수많은 인력과 자금을 투자, 82년 부터 지방지로는 꽤 권위있는 신문기사 DB를 구축했던 것이다.

84년 8월 부터 본격 서비스를 개시했고 中日는 정치, 경제, 사회, 문화, 국제등 대상 분야를 신문 기사 전 분야로 하고 있다. 중부 지방의 지방 뉴스도 대상, 중부 경제신문에 대해서는 중소기업 동향조사를 수록하고 있다. 수록 건수는 약 6 만건. 정보형태는 100 자 이내의 기사요약과 일부 기사에는 500 자 이내의 초록이 있고 검색 키워드는 자체 디소러스. 기업명, 인명, 단체명의 키워드도 포함하고 있다.

80년 2월 부터 상용 서비스를 시작하였고 서

비스 시스템은 ABC, 데이터베이스 명칭은 유니치로 명명하였다.

* 朝日新聞

80년대 초부터 DB개발에 착수, 84년 8월부터 자료입력에 들어가 85년 11월 自社 DB운영 체제를 갖추고 86년 4월 부터 상용 서비스를 시작했다. 朝日의 스포츠면을 제외한 전 기사를 수록하고 있으며 사진과 도표는 이미지가 나타나지 않고 유무 표시만 돼있다. 검색과정은 주제, 국별, 기사종류별 어느 것으로도 논리 검색이 가능하며 수록건수는 약 15 만건. 데이터베이스

은 각종 네트워크를 통해 서비스 하고 있다. 일본의 신문기사 DB 일람은 다음과 같다.

국내의 경우 중앙일보가 문헌정보, 인물정보 DB를 각각 3 만건, 1만 5천건을 입력, 사내서비스 하고 있고 미니컴 NES S-150/78 컴퓨터를 88년에 구입해 호스트로 삼고 있다. 중앙의 OB명은 "JOINS", 공급기관은 없으나 삼성태이타시스템에서 자체 망을 이용할것을 권유하고 있다.

한국경제는 뉴스플래시인 "한국케텔"을 86

일본의 신문 기사 데이터뱅크 일람표

OB명	서비스기관	수 록 대 상 지	수 록기 간	BD 작 성기 관	비 고
NIKKEI	NEED-IR	日經, 日經産業, 日經流通 등 日經발행 신문잡지	최근 4년	日本經濟新聞社	電氣의 DEM OS 이용
IEE	上 同	Coal News 등 에너지 관련 신문, 잡지	上 同	일본에너지 경제연구소	上 同
日經 News Telecom	上 同	日經, 日經産業 日經流通 등 日經발행 신문잡지	上 同	일본경제신문사	全文수록
NK-MEDIA	JOIS (일본과학 기술정보센터)	日刊工業新聞의 신기술, 신 제품 정보와 기업동향	83년 4월 이후	日刊工業新聞社	
TECHNO SEARCH	HINET	日刊工業, 日本工業, 化學工業 日報 電波通信 등 기술계 신문기사	81년 1월 이후	平和정보센터	
CHUNICHI	ACE	中日新聞기사, 中部經濟新聞 기사	84년 1월 이후	中日新聞社	
ASK	HINET	朝日新聞전체기사(스포츠면 제외)	85년 1월 이후	朝日新聞社	全文수록
YOMIDAS		讀賣新聞기사	86년 9월 이후	讀賣新聞社	全文수록
EL	ELNET	일본의 신문중 주요 신문기사	88년 12월 현재입력 중	Electronic Library Inc.	Scan image化

이름은 ASK이고 평화정보센터의 HINET를 통해 서비스 하고 있다.

* 其他

이밖에도 일본 에너지 경제연구소의 IEE, Electronic Library사의 EL, 일간 공업신문사의 NK-MEDIA, 평화 정보센터의 TECHNO SEARCH, 讀賣신문사의 YOMIDAS 등이 신문 혹은 잡지 기사를 데이터베이스에 수록, 전전 혹은

년부터 개발, 보급하고 있으며 데이터통신의 통신 망을 이용, 89년대 전국적으로 확산예정 이다.

수록 정보는 경제정보, 증권정보, 생활경제정보 등이 포함되고 있다. 매일경제는 닛케이의 각종 DB를 데이터통신 망을 이용, 회원제로 정보를 공급하고 있으나 사용언어가 일어와 영어뿐이라 공급이 부진한 편이다. 동아일보는 89년부터 본격 DB계획을 추진할 예정이며 이밖에 한국일보,

조선일보 등이 DB계획을 잠정적으로 추진중에 있다.

국내사의 경우 90년대에 접어들어야 DB가 본격화 될 전망이고 실제 활용시기는 93년경이후로 예상된다.

7. 결 론

신문 기사 정보가 DB화 함으로써 부가가치가 높아짐은 두말할 나위 없는 사실이다. 그러나 수작업에서 전산체제로 전환하기에는 여러가지 제약이 있게 마련이다. 기술인력, 예산, 인식 제고 등 해결과제가 많지만 모방 부터 시작, 차례 차례 개발에 착수한다면 그렇게 지난한 것만은 아니다. 최근 국가차원으로 전산망이 확산되는 가운데 뉴스DB도 거론되고 있으며 한국언론연구원이나 중앙일보 등이 기사DB를 착수했거나 추진중에 있는 것으로 안다. 실무자 입장에서 볼때 기사DB는 이용의 극대화, 상품가치의 고수준을 겨냥, 서둘러 착수해야 할 것으로 본다.

** 참고자료 1 : 국내 언론사 전산화과정

국내 언론사에서 처음으로 전산부를 발족시킨 회사는 중앙일보로 75년에 발족, 79년에 IBM 370을 도입해 본격적인 전산화를 도모했다. 중앙일보는 同 기종으로 중앙일보 IR 디소리스와 기초적인 업무전산에 속하는 급여 업무등을 처리했다. 이후 용량 부족을 절감하고 81년 8월 FACOM M340S로 교체하면서 잡지목록 생산과 인사업무를 시작한 다음 판매, 광고등으로 영역을 확대하게 되었다. 업무는 늘어나고 FACOM 기종의 Overload가 계속되었다. 그러자 88년에 접어들어 M760/4로 기종을 교체할 계획을 진행 시키고 있다.

두번째는 한국일보로 79년 2월에 발족했고 80

년 10월 NCR1-8200 마이크로급 컴퓨터를 도입한 후 85년 12월 VAX750을 추가 도입했다. 한국일보는 同기종을 급여, 주간지 MAILING, 보급소관리, 확장사원 관리등에 사용하고 있는데 이와 별도로 CTS도 차근차근 진행시켜 나가고 있다.

세번째는 경향신문이며 81년 5월에 발족, MV 4000 기종을 83년 12월에 도입했고 현재 판매, 광고, 인사, 경리, 자재 관리, 출판판매, 출판광고 등 각종 업무 전산화를 완료한 상태다.

그 다음은 서울신문, 82년 4월 발족, 82년 10월 HITACHI 325 마이크로급 컴퓨터 도입, 급여, 일반회계, 보급소관리, DBM등을 주력하고 있으며 조선일보는 84년 9월에 전산부가 발족해 85년 5월 BURROUGHS-B1990 기종을 도입, 판매 급여, 출판, 자재관리, 증권시세, 광고업무등을 완료했다.

이들 회사는 대부분 1개국 단위 업무의 전산화를 완성하는데 시일이 9개월정도 소요된 것으로 나타났으며 급여관리는 3개월 가량 시간이 걸린것으로 알려졌다.

동아일보는 84년 3월 전산부가 발족, 고대 IBM4381 기종을 리모트로 불러 급여관리, 광고, 판매, 출판판매 등 업무를 87년까지 완성한 후 인사, 자재관리등을 88년내 추진중이다. 또 83년 10월 IBM9735/40 기종을 도입, 자체 컴퓨터를 보유하며 일반 관리업무를 확충시켜 나가고 있다.

** 참고자료 2 : 언론공용 DB

국가기간전산망 계획중 데이터베이스 구축 분야에서 프로젝트명 P2007은 언론 공동 데이터베이스다. P2007은 현재 언론연구원에서 진행 중인 언론사 공동 데이터베이스 구축사업을 기술적, 재정적으로 지원해 주는 사업으로 필요한

경우에는 제도적 뒷받침(예를들어 法的지원)도 할 수 있다고 돼있다.

언론연구원은 87년중 대충의 언론 공용 DB를 레이아웃한후 전산 관계자 언론관계자들을 초치, 몇차례 자문회의를 가진 다음 신문사 사주들의 동의를 구하지 않은채 본격 추진 계획을 확립했다. 88년 언론연구원은 언론 各社 사장들을 초청해 파일럿시스템을 견학시켰고 곧이어 정보관련부서인 조사부와 과학부 기자들을 불러 동 시스템을 시연했다. 前者는 일부 긍정 반응을 보였으나 後者에겐 호된 비판과 함께 질책을 당했다. 후자들은 파일럿시스템을 구축한 KO-PEC측과 주관자인 언론연구원의 DB구축 실력과 의지를 질타 했던 것. 후자들은 특히 정보를 생성하는 각 신문사들이 생산자가 되고 언론연구원이 배급자가 되는 시스템을 제시하는가 하던 신문이란 속성을 아는 실무자가 단 한사람도 DB구축에 참여하지 않아 생긴 에러를 낱날이 지적하기도 했다.

즉 신문기사의 속성과 이용관행을 고려하지 않은 점외에 한자처리, 제목처리, 맞춤법문제, 주제어조절, 디소러스채택여부등 조차 갖추지 못한 상태여서 실무자들로서는 공용 DB를 의심치 않을수 없었던 것. 이점은 언론연구원측도 잘못을 인정하고 시정을 약속한바 있다.

언론연구원은 파일럿시스템이 이같이 부실했던점을 * 자체인력 부족 * 신문기사 분석 미비 * 시스템 미설치 * 제도적인 뒷받침 부족 등으로 분석하고 관련자들을 초치, 보완작업에 착수해 단계적으로 완벽을 기할 예정이라고 설명하고 있다. 88년 12월중 HP 3000/950기종을 들여와 89년부터 입력 작업을 개시할 예정으로 있으나 파일럿 시스템 시연장에서 지적된 문제점이 해결되었는지, 실제 이용자들이 제시한 입

력 양식 변경이 이루어 졌는지 발표가 없어 확실히 알 수 없으나 내부 관련자에 따르면 큰 변경이 없다하여 문제가 되고 있다. 또 언론 공용 DB관계 인력은 DB와 거의 관련이 없었던 관리자 3인, 전산인력 5인 등이 고작일뿐 언론계 실무자는 아직도 참가하지 않은한 상태여서 과연 원만한 운영이 될지 미지수다.

그러나 과감히 DB사업에 손을 대 언론사에 정보를 POOL 하겠다는 의지는 대단하다. 또한 언론계 공용DB의 효시로 일반인들의 관심과 기대 역시 대단한 편이다. 그럼에도 불구하고 데이터베이스 유통과정상 제작업자(Producer) 분배업자(Distributor) 전송업자(Data Carrier)의 모든 역할을 다하겠다는 것은 의심할 여지가 많다. 왜냐하면 언론연구원은 신문을 찍지않는 연구기관이라 Producer도 아니고 중앙지 9종의 신문기사 전체를 입력, 검색할 수 있는 방대한 재원을 가진 Distributor도 아니다. 또 회선을 보유한 Carrier도 아니다. 어느 하나도 갖추지 못한 언론연구원이 만약 저작권료를 요구할 수 있는 Producer 격인 신문사에서 그 사용료를 요구하면 어떻게 할지, 실제 Carrier입장인 DACOM과 전기통신공사가 상당한 회선료를 요구하면 어떻게 할지 그 대응태도가 주목된다.

언론공용 DB는 국가적인 사업이다. 국가적인 사업은 언론 DB전문가나 실무자 하나없는 언론연구원이 독자적으로 수행할 수가 없다. 그러므로 언론연구원은 중앙지최초의 신문사와 언론DB전문가, 실무자, 관련 학자들을 포함한 위원회에 그업무를 맡기고 Distributor역만 해내도 큰 공헌을 할수 있을 것이다.

* 참고자료 3

기 간	DB제작현황 및 DB내용, 종류	비 고
산 업 연 구 원	<ul style="list-style-type: none"> · 국내 관련 DB 7종 · 해외 DB 구입 보유 17종 	<ul style="list-style-type: none"> · 자체제작 및 용역 DB 제작 · 해외 DB 구입 보유
국 회 도 서 관	<ul style="list-style-type: none"> · 정기간행물 기사색인 · 국내 식·박사학위 논문목록 	· 색인지발간 각 도서관 배포
서 울 대 도 서 관	<ul style="list-style-type: none"> · 장서목록 (영문) 	· 진행중
한국전자통신연구소 (ETRI)	<ul style="list-style-type: none"> · 전기통신관련 DB 	· 진행중
한 국 전 력	<ul style="list-style-type: none"> · 전력관련 DB 	· 진행중
코 오 룡	<ul style="list-style-type: none"> · 섬유, 화학 관련 DB 	· 자체사용중
전 기 통 신 공 사	<ul style="list-style-type: none"> · 전화번호부 DB 	· 자체사용중
경 제 기 획 원	<ul style="list-style-type: none"> · 경제통계 DB 	· 자체사용중
출 판 문 화 협 회	<ul style="list-style-type: none"> · 국내출판물 총목록 	· 용역제작
정 신 문 화 연 구 원	<ul style="list-style-type: none"> · 한국학연구 인명록 	· 자체제작
관 세 청	<ul style="list-style-type: none"> · 수출입관련 통계 DB 	· 자체제작
특 허 청	<ul style="list-style-type: none"> · 특허, 의장, 상표 DB 	· 자체제작
매 일 경 제 신 문	<ul style="list-style-type: none"> · 기업정보 DB · 닷케이 DB 정보 	<ul style="list-style-type: none"> · 계획중 · 닷케이대행
한국인구보건연구원	<ul style="list-style-type: none"> · 현재 미제작 · 인구, 가족계획, 보건 등의 자료를 입력할 계획 	· 계획중
중 양 기 상 대	<ul style="list-style-type: none"> · file형태로 보관중임 (M/T) · 국내기후자료 ('84년~'85년) 	· 구축
동 양 나 이 론 (주) 연구 소	<ul style="list-style-type: none"> · 섬유, PLASTIC 관련 정보를 취합하여 입력중 	· 자체 활용중
농 촌 진 흥 청	<ul style="list-style-type: none"> · 국내 농업학술잡지, FAO의 AGRIS DB보유 	· M/T보유
해 외 건 설 협 회	<ul style="list-style-type: none"> · 해외건설 관련 11개분야 DB 제작 · 해외진출 건설사로부터 주별 주기적으로 공식 접수되는 모든 해외관련 건설자료를 입력하여 건설부에 보고하고 각 회원사에 게 공개 	· 자체제작중
한 국 무 역 협 회	<ul style="list-style-type: none"> · 수출입, 무역관련정보의 통계 DB 구축중 · 상공부에서 공개를 허용한 DATA에 한해서 책자 발간 	· 관련기업체 활용
대 한 무 역 진 흥 공 사	<ul style="list-style-type: none"> · INQUIRY 정보, 조사정보, 지역정보, 전시회 개최정보 등의 DB를 구축중 	· 관련기업체 활용
국 토 개 발 연 구 원	<ul style="list-style-type: none"> · 국토정보 DB 완료 및 UPDATING (인구, 고용, 경제, 산업, 토지, 자원 및 에너지, 교통, 통신, 주택, 기반시설, 사회복지, 자연조건) 	· 자체이용
조 달 청	<ul style="list-style-type: none"> · 주요기관, 업체 관련 내용 	· 자체 File구축

기 간	DB제작현황 및 DB내용, 종류	비 고
한국에너지연구소	· 에너지통계분야, 기술분야의 DB 구축	· 자체이용
대 법 원	· 대법원판례 DB입력중 · '80년도 이전의 것은 색인 형태로, 그 이후 것은 전문수록중	· 진행중
KAIST SEC 자동화시스템연구부	· 시사영어사 사전DB화	· 구축
국립중앙도서관	· 81년 이후의 납본된 단행본과 83년 이후의 석·박사학위논문 입력중	· 구축
한국개발연구원	· Macro DB, Micro DB, OECD, IMF, 국내외 경제동향 DB입력중	· 구축
한국학술진흥재단	· 외국 학술잡지 목록 및 국내 교수 명록을 KIET에 공익 의뢰 입력중 · 향후 대학부설연구소 총람, 학회총람, 해외유학생 인명록, 국내 외국인유학생 인명록을 DB화 예정	· 구축 및 진행중
한국건설기술연구원	· 건축, 토목, 기계, 전기등 4 분야에 관한 각종 공사 시방공법 및 기술에 관한 내용을 DB화하여 영문으로 초록작성중	· 자체제작중
한국표준연구소	· 국가별, 단체별 표준 ISO의 54종 규격의 DB구축	· 자체이용
한국동력자원연구소	· 에너지 통계 DB구축중 · 신재생에너지 Data입수	· 자체이용
터 키 연구 소	· PC를 이용하여 기수정보초록 수록	· 자체이용
종 로 서 적	· 일부 외서에 대한 서지 사항을 입력시켜 재고관리용으로 활용중	· 자체이용
한 국 경 제 신 문	· 경제뉴스플래시 · 증권정보	· 회원서비스 예정
교 보 문 고	· 출판사 MASTER DB, 도서별 MASTER DB를 구축중	· 자체이용
중 앙 일 보	· 국내 인물정보 FILE 인물사진 18만장 M/F화	· 자체이용
한 국 증 권 전 산 (주)	· 증권종합정보(상장기업체 증권정보) · 각국 환율정보	· 회원제 서어비스중
증 권 감 독 인	· 상장기업체기업신용정보	· 계획

參 考 文 獻

- 1) 石坂悅男, 매스미디어 산업의 전환(정보혁명으로 어떻게 변할 것인가) 有斐閣, 1987.
- 2) 南相石, 전산화 무엇을 어떻게 할 것인가, 東友, 1987.6월호 pp.22-30
- 3) 李寬基, 정보화 시대의 신문기사정보은행, 우정출판사, 1986.
- 4) 서울대 사회과학연구소, 정보화사회 도전과 대응, 서울대학교출판부, 1986.
- 5) 金尙鉉, 한국매스미디어사회, 나남, 1986.
- 6) 吳海石, 데이터베이스, 정익사, 1986.
- 7) 盧仲鎬, 종합경영정보시스템, 정민사, 1982.
- 8) 電算網調整委員會, 國家電算總覽, 전산망조정위원회, 1988.
- 9) Wilfrid Lancaster 저, 김태승. 운구호 공역, 정보검색시스템, 구미무역, 1985.
- 10) 李英子, 李慶浩, 정보학개론, 경북대학교출판부, 1987.
- 11) Cay Weitzman, Distributed Micro /Minicomputer Systems, Prentice-Hall Inc., 1980.
- 12) 金東訓, 정보화 사회 가능성과 문제점, 정우사, 1987.
- 13) 南相石, 사진자료전산화, 사진기자 1987 겨울호, pp.111-120.
- 14) 이상헌, CD-ROM 기술을 이용한 도서관 편목시스템 개발에 관한 연구, 연세대 석사학위논문, 1987.
- 15) 노정순, 해외언론사의 데이터베이스현황, 신문과 방송, 1987. 통권 199 호.