

멀티미디어로 DB 서비스로 구현되는 Information Service

REUTER Korea Inc.

BIP/Customer Liaison Executive Eun Sun Kang

로이터의 데이터 베이스의 역사

1980년 FINSBURY DATA는 수많은 source로 부터 뉴스를 모을수 있는 시스템을 개발하여 이것을 공문서(archive)로 보관로 하여 이것을 HP dumb terminal과 모뎀을 이용하여 search가 능하게 만들었다.

많은 source 확보하였고 사용자가 늘자 다른 상품으로 개발하였다. 휘스버리가 개발한 한 상품은 Finsbury stand-alone 터미널을 통하여 12개월주기의 데이터 베이스의 한종류를 제공하였고 이는 회사의 구체적인 뉴스를 검색할수 있게 하였다. 이 상품은 회사별로 검색할수 있게 하였는데 이를 COMPANY NEWS YEAR이라 불렀다. 이 상품은 현재도 계속 유지되고 있었으나 1997년 6월부터 사용할수 없으며 이는 IDN에 연결된 게이트 웨이를 이용하여 로이터 터미널을 통하여 접속할수 있다.

로이터가 Finsbury data를 구입하여 이를 로이터 파일로 이름을 변경하였으며 이 textline은 그 당시에 고객층이 얹었다. HP 터미널 인터페이스는 사용하기에 어려웠고 단지 회사내의 librian이나 research을 중심이었다. 복잡한 검색은 단지 장시간동안 감도가 좋은 라인으로 접속하여 겨우 이루어졌다. 전용 터미널 설치는 고객의 PC에서 이물레이션이 작동할수 있게 하였으나 이것은 검색시간의 향상을 주지도 그리고 복잡한 검색에도 영향을 주지 못했다.

이시기에 Historical text product의 주 수입원은 제3의 회사에게 로이터가 수집한 데이터를 판매함으로써 얻을수 있었다. Mead, Maid and FT Profile와 같은 상대 회사였고 이들은 자체의 데이터 베이스 상품을 로이터의 데이터를 부수적으로 이용하면서 제공하였다.

1990년경 Historical Text Development Group은 두가지에 노력을 기울이기 시작하였다. 첫째는 마이크로 소프트 윈도를 통하여 향상된 사용자의 인터 페이스로 사용자가 Topics, Sources, Countries, Markets등을 간단하게 검색할수 있게 하는 것이었다. 이것이 오늘날의 전조인 로이터 비즈니스 브리핑이다.

두번째 목표는 검색시간을 단축하는 것으로 데이터 베이스의 규모와 고객층이 확대되면서 검색시간이 너무 느리게 때문에 유용한 상품으로의 가치가 하락되었다. 해결방법은 AMT(Distributed Memory Technology)라는 회사가 개발한 hardware 인 DAP(distributed Arrayed Processor)이었다. DAP 는 상당히 간단하며 수천열의 CPU 와 많은 메가 바이트의 빠른 Static Ram 이었다. 이것은 index 와 headline data 를 메모리에 저장하고 사용자가 textline 을 통하여 검색을 하면 DAP 가 이에 해당하는 headline 데이터의 명단을 매칭하는 방식으로 host 에 접속하여 소모하는 시간을 단축하는 방법이었다. 이때까지만 해도 DAP 는 단지 대학이나 연구소등에서만 사용되었다.

DAP 와 윈도 인터페이스의 개발은 1992 년말경에 완성되었다.

빠른 검색은 'point and click'의 인터페이스의 유용성을 의미하였으며 단지 researcher 와 librarian 에게만 국한된 상품을 메니저와 직원들에게도 제공할수 있는 기회였다.

이로인해 로이터의 비즈니스 브리핑(Reuters Business Briefing)이 첫 선을 보여졌으며 로이터의 비즈니스 인포메이션 상품(Business Information Product)의 개념이 처음 탄생한것이다.

이후에 이는 여러 언어의 중국어와 일본어 버전의 RBB 을 아시아에 소개하게 되었다.

최근에는 웹 기술에 의한 상품류를 소개하고 있다. 현재 RBB 는 인터넷이나 로이터의 웹으로 접속가능하며 이런 경향으로 새상품들이 소개될 전망이다.

현재의 2.0 버전은 획기적인 변화의 인터페이스와 부가적인 down-line loadable 소프트웨어의 update 및 다국어 이용이 가능하며 acrobat 를 통한 retrieve 도 가능하다.

현재 RBB 를 통한 멀티미디어의 전달도 가능하리라 보며 그래픽 뿐만아니라 비디오 클립핑 등 앞으로 무한 하리라 보여진다.

데이터가 어디에서 오는가?

1980 년이래로 데이터 베이스를 전송하여주는 출처(source)는 늘어나고 있으며 앞으로도 그럴것이다. 현재 800 개사의 다른 source 로 부터 데이터를 전송받고 있으며 이는 약 3000 여개의 출판물로 Guardian, American Banker, Times 등이다.

현재 로이터는 여러 방식으로 데이터를 전송 받는데 주된 방식은 HPSN(로이터의 네트워크)이다. 데이터 제공자는 지역 HPSN 패드에 전화를 걸어 데이터를 host 에 download 한다. 주 편집 cluster 에 다운로드 하는 것을 쉽게 하기위하여 FONT END CLUSTER 가 소개 되었고 거의 대부분의 downloading 이 이를 통하여 이루어 진다.

HPSN 뿐만아니라 데이터 제공자들은 PPS(public packet switching network)로 직접 연결할수 있

다. 물론 정기적으로 테이프나 디스크로도 공급받는다.

영국에는 5개의 주 제공자들이 개별라인 접속을 통하여 로이터의 많은 량의 데이터를 하루 종일 전송하고 있다.

또다른 전송 메커니즘은 다른 아닌 Wire Editing System(WES)으로 아직 source processing 으러 이전되지 못한 전송자들을 위하여 사용하고 있으나 이는 1997년중반경에는 모두 source processing 시스템으로 변경될것이다.

데이터 코딩 작업

많은 source 를 topic, industry 및 country 별로의 검색을 기본화 하기 위해서는 로이터내의 자체 기본 코딩시스템이 있다.

이전에 언급한 source processing 은 많은 데이터 포맷을 한가지의 내부 기본 포맷으로 변형한다. 이 포맷은 필요한 코드가 자리 잡아야 할 위치를 제공한다.

TIS(Topic Identification System) 이라 알려진 5 micro VAX 가 LAN 에 있다. 이는 인공적 지식 기기(artificial intelligence application)이라 불리는 'construe'로 LUCID LISP 용어로 작성된다. construe 는 최대한 필요한 코드를 문서를 분석하는데 가능한 rule 을 사용하여 만들어 내는 것이다.

'source processing'에는 두개의 TIS box 가 정해지고 WES 에는 3개의 TIS box 가 주어진다. 최종적으로 모든 데이터는 편집자들에 의하여 검토되거나 삭제 그리고 필요한 코딩을 한다.

Wire 로 전달되는 데이터는 모두 실시간 데이터이므로 모두 실시간으로 전달되어야만 유용한 자료가 되기에 이는 시간이 상당히 중요하다.

주 편집기능을 다 통과한후에 모든 source processed data 와 wire data 는 TOT file 에서 종결된다. TOT file 은 일반적으로 하루안의 가공된 한 source 에 의한 데이터로 이루어진다.

TOT 파일은 어디로?

TOT 파일은 시스템의 두곳으로 보내진다. DAP 데이터 베이스와 데이터 베이스 제공하는 제 3의 계약자이다. TOT 파일은 DAP 데이터 베이스 의 하루 분량의 파일에 추가가 되어

DAP 가 색인화 할수 있도록 한다.

TOT 파일이 도달하는 다른곳은 RHTF text feed cluster 이다. 이는 TOTO 가 4-5 일 머무는 속으로 이는 리그식의 데이터 베이스이다. 각각의 제 3 의 데이터 베이스 분배자가 원하는 것을 'pointer'라는 도구를 통하여 download 할수 있다. 만일 금요일에 download 를 중지하였다면 pointer 는 월요일에 다시 download 할 때 어디에서 download 하여야 하는지 기억하여 시행한다.

TEXTLINE DISK BANK

텍스트라인 호스트는 약 약 750 GBYTE 의 디스크 저장소로 Status, DAP, Rapid News 그리고 부가적인 데이터베이스를 보관하는곳이다. 현재 한주의 약 1 gigabyte 정도씩 추가 된다.

데이터로의 접속

현재 모든 사용자는 로이터 비즈니스 브리핑 방식의 인터페이스로 접속한다.

저장된 오랜 데이터에 접속하기 위해서는 하나의 예로 1994 년 5 월부터 1996 년 8 월사이의 데이터라면 검색은 처음에 DAP 1 을 먼저 검색하고 그다음 전체 DAP 를 검색할것이다. 1996 년 8 월 이후의 검색은 DAP 2 에서 진행될것이다.

만일 1994 년 5 월이전의 데이터를 검색하려면 먼저 DAP 를 그리고 다시 Status 시스템으로 이동할것이다. 이는 Status 시스템의 검색은 DAP 데이터 검색보다 상당히 느리기 때문이다.

최근에야 비로소 다기능 검색이 가능하여졌다. 이는 Status 데이터 베이스가 다루는 날짜의 데이터를 DAP 로도 다룰수 있게 하기 위해서이다. 평가한 결과 1994 년 5 월 이후의 데이터는 거의 검색되어 지지 않으므로 Status 에 저장하는 것은 하나의 소모이기 때문이다.