

우리말 데이터베이스의 정보검색

한국사람은 너무나 일찍 망각하여 버리는 악습이 있어서인지 오늘날에 와서는 한글부호에 관한 이견을 말하거나 생각하는 사람조차 몇 안되는 것처럼 보인다. 우리말 데이터베이스 산업을 위해서는 한글부호의 표준화에 이어서 정보처리용 국어의 표준화로 이어져서 '띄어쓰기'의 어려움을 어떻게 극복할 것인가에 이르기까지 온 한글부호 문제가 제기되었을 때만큼의 관심이 지속되어야 하지 않을까, 아쉬운 생각을 금할 수가 없다.

영 어 단어 몇개를 알고 또 정보검색논리 몇개를 알면 그만 인터넷 전문가가 되어 버리고, 정보검색이라면 영어로 만들어진 데이터베이스를 주로 찾는 기술로 알려지고 있다. 간혹 우리말로 된 데이터베이스가 있긴 해도 영어로 찾는 데이터베이스에 비해서 색인기술이나 초록기술이 영어의 경우처럼 세밀하지 못해 데이터베이스 축에 끼어넣지도 않고 있는듯 하다.

왜 우리말로 된 영어로 된 데이터베이스와 같은 것들을 만들지 못하고 있는 것일까? 우선 영어 단어의 [자동색인] 기술을 우리말 데이터베이스에 적용하지 못하기 때문일 것이다. 영문 데이터베이스안에 들어있는 모든 데이터는 자동적으로 유의적인 단어를 추출해서 색인하기가 용이하다. 왜냐하면 영문은 모두 [띄어쓰기]로 쓰여져 있기 때문이다.

이에 비해서 일본어 데이터베이스인 경우에는 일본어는 [띄어쓰기]가 없고 모두 붙여쓰기 때문에 입력전에 인위적으로 유의적인 단어식별을 위한 식별기호를 매단어의 앞뒤에 부가하고 있다. 일본어처리가 영어보다 그래서 더 어렵다고는 하지만, 식별기호를 붙이는 방법이 표준화되어 있어서 데이터베이스 제작에는 큰 문제가 없는것 같다.

그러나 우리말인 경우는 어떠한가? 이런 말이 있는지 없는지는 몰라도 영문과 같은 '띄어쓰기'

도 아니고 일문처럼 '붙여쓰기'도 아닌 '반띄어쓰기'이기 때문에 '자동색인'도 불가능하고 또 인위적인 식별기호 부가도 참으로 어렵다. 그래서 대체로 이 두가지 색인방식을 포기하고 데이터베이스를 만들어 버리기 때문에 영성한 것이 되어버리고 만다.

'과학기술'이란 말이 있다고 하자. 이것이 '과학과 기술'을 의미하는 것인지 '과학적 기술'을 의미하는 것인지 경우에 따라서 붙여써야 하기도 하고 띄어써야 하기도 하는가 본데, 일반인은 이것을 분명히 구별해서 사용하지 못하고 만다. 그래서 이른바 붙여쓴 '복합용어'가 많아지고 있는 추세에 있다. 그런데 국어문법상에도 '붙여써도 좋고 띄어써도 좋은' 애매한 문법이 너무나 많아서 뚜렷한 '띄어쓰기' 규칙이 없다는 것이 자동색인을 제대로 할 수 없는 원인이 되고 있다.

국민들이 가장 널리 읽고 있는 글이라면 역시 신문이다. 신문의 기사에서도 띄어쓰기가 국어문법대로 되지 못하고 있다고 나무라는 것은 아니다. 현행띄어쓰기 문법으로서는 데이터베이스 제작을 위한 자동색인을 할 수 없다는 것이다. 용기를 내어서 대폭적으로 붙여쓰기를 허용한다면 데이터베이스에는 편리할지 모르지만 너무 붙여쓰다가는 '국가적재정손실대책'이란 새로운 긴 복합단어가 일상화되어 버린다는 것이다.

몇년전에 온나라를 떠들썩하게 만들었던 '한글의 컴퓨터부호'가 풀어쓰기 부호(조합형)로 할 것인가 모아쓰기 부호(완성형)로 할 것인가에 대하여 논란이 있었던 것을 기억하고 있을 것이다. 이것이 결국은 모두가 필요한 것인데도 불구하고 대립된 것처럼 인식되어서 국론이 분열된 것처럼 언론에서 떠들어 댔다. 이렇듯 관심을 가진 것이 단순한 '국어사랑' 이라기 보다 '국어정보의 컴퓨터 처리' 내지는 '우리말 데이터베이스의 과학적인 제작'을 위해서 논의가 계속되어야 했었다.

한국사람은 너무나 일찍 망각하여 버리는 악습이 있어서인지 오늘날에 와서는 한글부호에 관한 이견을 말하거나 생각하는 사람조차 몇 안되는 것처럼 보인다. 우리말 데이터베이스 산업을 위해서는 한글부호의 표준화에 이어서 정보처리용 국어의 표준화로 이어져서 '띄어쓰기'의 어려움을 어떻게 극복할 것인가에 이르기까지 온 한글부호 문제가 제기되었을 때만큼의 관심이 지속되었어야 되지 않을까, 아쉬운 생각을 금할 수가 없다.

필자가 예를 들었던 몇가지의 정보검색 사례나 최근에 교육방송에서 온 국민이 모두가 알아야 하는 것처럼 강조하고 있는 '인터넷 정보사냥'도 모두가 영어의 정보검색이다. 그래서 영어문법을 우리말 문법보다 더 자세히 알아야 하고 우리말 정보의 검색은 하지않아도 되는것인지 안타깝기 이를 데 없다. 과연 이것이 세계화인가? 세계화에서 더 중요한 것은 우리말 데이터베이스의 세계화가 아닐까?

사람들은 다음 단계를, 영어를 우리말로 자동번역하는 것이라고 한다. 이 말도 맞는 말이지만 사실은 우리말을 영어로 자동번역하는 것도 이에 못지않게 중요한데 이것을 그 다음의 단계로 보고 있는 것은 영어에 대한 종속성을 더욱 크게 할 뿐이라고 본다.

필자는 고령자들의 PC통신 공동체인 '원로방'이라는 단체를 운영하고 있다. 노인이란 말은 노인들에게 대단히 듣기가 거북한 말이다. 일본도 미국도 마찬가지이다. 그래서 그런지 미국에서는 Senior Citizen이라는 덜 거북한 용어를 창안해

내어서 쓰고 있으며 PC통신망의 이름도 SeniorNet란 이름을 붙일 정도이다.

일본에서는 원숙세대란 의미로 영어의 Mellow Generation이란 일본제 영어를 만들어서 쓰고 있으며 PC통신망도 MellowNet란 용어를 쓰고 있다. 우리는 고령자란 말도 적합할 것 같지 않아서 '앞선이'란 용어가 어떤가라고 제안중에 있다.

언제부터인가 '실버산업'이란 말을 자주 읽고 듣게 되었다. 신문에 일단 실버산업이라고 나니 이것이 벌써 표준말이 되어버렸다. 그래서 그러려니하고 미국의 실버산업은 어떤가를 정보검색하여 보기로 마음을 먹었다. 이것을 굳이 우리말로 은빛 산업이니라고 억지로 번역해서 쓰면서...영어로 이것을 옮겨 쓴다면 단연코 Silver Industry임에 틀림이 없다. 이것을 정보검색을 위한 Keyword로 삼아서 다이알로그(Dialog)에 담겨진 몇가지의 데이터베이스에서 검색하여 보았다.

우리가 생각하던 실버산업에 관련되는 정보는 하나도 검색되지 않았다. 검색된 정보들은 모두가 '은광산'에 관한 정보뿐이었다. 은을 캐내는 광산업을 일컬어서 실버산업이라고 하고 있다. 그러면 그렇지 노년기를 표현하는 실버란 용어는 일본식 영어임에 불과하였다. 일본의 전철을 타보면 '실버시트'라고 특별히 표시된 곳에는 '노약자 좌석'이라는걸 굳이 영어화해서 그것의 일본식 발음을 마치 영어의 외래어인양 쓰고 있었다는 사실을 알았다.

우리나라에도 실버용품점, 실버타운, 실버산업, 실버식품 등 실버란 용어가 범람하고 있다. 이제는 일본식 영어라고 거부하기조차 어려울 정도로 일반화가 되어 버린듯 하다. 우리말 데이터베이스를 어서 실용화하는 환경을 조속히 만들어야 하는데 오늘도 영어 데이터베이스의 흉내만 내는 세계화를 위해서만 노력하고 있는 것 같다.

〈(재) 한국정보유통센터 회장〉