



# KISTI와 함께 국가대표 시소러스를 만든다, 오롬정보

글 · 사진 \_ 김희정 · 대외협력팀 · khj@kisti.re.kr

시소러스, 온톨로지... 일반인에게는 너무나 낯선 단어들이다.  
 그러나 숨을 쉬듯 자연스러운 일상이 된 인터넷 검색을 통해 우리는 이 용어들을 항상 접하고 산다.  
 시소러스(thesaurus)란 기계검색에 활용하기 위해 유의어, 반의어 등을 포함한 용어의 상하위 개념 관계를 나타낸  
 일종의 색인사전을, 그리고 온톨로지(ontology)란 지식을 효과적으로 표현하기 위해  
 정보의 의미를 부여하고 정보 간의 관계를 설정하여 표준화된 지식의 표현, 공유 및 재사용을 가능하게 하는 기술을 말한다.  
 그리고 국내에 구축돼 있는 시소러스의 약 70% 정도를 만든 기업이 바로  
 직원 40명의 벤처기업, (주)오롬정보다.



우리 연구원이 오롬정보와 본격적인 협업을 하게 된 것은 지난해 '지식기반정보유통플랫폼 개발연구사업'을 시작하면서부터다. 연구자들이 보다 활용도 높은 정보를 정확하게 찾을 수 있도록 최신 과학 기술용어에 대한 시소러스를 구축하지는 취지에서 이 사업을 추진하게 되었고, 최적의 업체를 찾던 중에 미개척분야인 시소러스를 이미 13년째 사업화하고 있는 오롬정보를 선택하게 된 것이다.

시소러스는 물론이고 인터넷이라는 용어조차 익숙하지 않던 지난 1993년부터 시소러스 구축사업을 시작한 오롬정보는 과연 어떤 기업일까?  
 서울 충정로2가에 위치한 오롬정보에서 느낀 첫 인상은 '짤다'는 것이었다. 기혼자보다 병역특례로 근무



하는 20대 초반의  
 옛된 청년들이 훨씬 많은, 그야말로  
 벤처 냄새가 물씬 풍기는 사무실. 그러나 오름은  
 18년이라는 결코 짧지 않은 역사를 자랑하는 기업이다.

“초창기 직원들 대부분은 문헌정보학과 출신이었습니다. 당시만 해도 도서관 전산화가 거의 안 되어 있었고, 기껏해야 일본에서 들여온 전산시스템을 주먹구구식으로 사용하던 때였습니다. 그렇게 후진적 상황이던 도서관을 우리 힘으로 멋지게 전산화해보자는 취지에서 오름정보를 세우게 됐죠. 덕분에 대학교의 문헌정보학 교재에 ‘오름’이라는 이름이 실릴 정도로 도서관전산화시스템 업계에서는 독보적인 존재가 되었습니다. 하지만 시간이 지날수록 정체된다는 느낌이 많았어요. 앞서나가기 위해서는 뭐든 한 발 먼저 시작해야한다는 생각을 하게 됐고, 그래서 새롭게 주력하게 된 것이 바로 시소러스 사업입니다.”

시소러스의 개념도 잡혀있지 않던 시절이었으니, 당연히 대부분의 사람들이 사업의 시장성과 비전에 의문을 품었었다고 한다. 그러나 오름은 시소러스의 가능성과 중요성을 확신하고 있었고 93년 국방과학연구소의 국방용어 구축을 시작으로, 법률·아동·디자인 등의 다양한 용어 그리고 2002년에는 국립중앙도서관과 함께 20만개의 시소러스를 구축하는 등 빠르게 시소러스 분야에서 입지를 넓혀왔다. 그러나 지난해부터 시작된 KISTI와의 협업은 오름에게 있어서 매우 신선한 ‘충격’이었다고 한다.

“시소러스 분야에 있어서는 나름대로 최고라는 생각을 갖고 있었습니다. 그런데 KISTI 분들과 일하면서 ‘우리가 많이 부족했구나’ 하는 생각을 정말 많이 했어요. 사실 시소러스는 언어학이 배경이기 때문에 상당히 주관적일 수밖에 없습니다. 정확하고 객관적인 기준을 만드는 것은 정말이지 어려운 일이에요. 그런데 KISTI는 모든 시소러스를 명확하고, 합리적이고, 논리적으로 증명하려 했고, 누구나 객관적으로 시소러스를 이해할 수 있도록 ‘시소러스 구축지침’을 만드는 데도 성공했습니다. 현재 버전 12까지 나와 있는 이 지침을 출판할 계획까지 잡고 있는 건 정말 대단한 자신감이라고 밖에는 말할 수가 없어요.”

2년째 KISTI와의 협업을 총괄하고 있는 오름정보의 이승준 팀장은 KISTI 덕분에 배운 것이 너무나 많다고 거듭 강조했다. 그는 특히 ‘시소러스 구축지침’의 출판계획에 대해 놀라움을 금치 못했다. 객관적인 시소러스 기준을 만들었다는 것만 해도 대단한 일인데, 이것을 출판까지 해서 모든 사람들 앞에 논리적으로 증명해 보이는 것은 결코 생각해 본 적이 없는 일이라는 것이다.

“처음 사업을 시작할 때 KISTI가 가장 먼저 요구한 것은 신문과 학술지 등에서 최근 사용되는 용어들을 추출해 시소러스를 만들자는 것이었습니다. 저희 입장에서는 당연히 반대를 했죠. 사전에 실려 있는 용어를 가지고 시소러스를 만들어도 평가가 분분한데, 충분히 검증되지 않은 신조어의 시소러스를 구축한다는

것은 정말이지 ‘모험’이라고 밖에는 할 수 없었거든요. 하지만 과학기술 용어의 생명주기가 평균 3년에 불과하다는 것을 고려하면, 사전에 실패 정도로 정착된 용어는 이미 활용도가 떨어진다는 것이 KISTI의 입장이었습니다. 연구자들의 R&D를 돕기 위해서는 가장 활용도 높은 최신용어의 시소러스를 구축해야만 한다는 거죠.”

KISTI도 오름정보도 모험이라는 것은 충분히 알고 있었지만, 일단 해보기로 결정을 내렸다. KISTI가 전문 용어자료추출기로 신문, 잡지, 최신 논문 등에서 용어를 추출하고 시소러스 구축지침을 만들면, 이것을 기준으로 오름정보가 시소러스를 구축하는 작업이 수십 번 반복됐다. 그리고 드디어 1년여 만에 최신 범용 과학기술 용어 1만 5천개에 대한 시소러스가 구축됐다. 용어 사이의 계층적 관계까지 객관적으로 체계화 한 온톨로지형 시소러스였다. 지난해 일단 범용용어 시소러스 구축에 성공한 KISTI는 올해는 IT, 다음해에는 BT, 그 다음해는 NT... 이런 식으로 주요 과학기술 분야 전문용어 시소러스를 구축해 나갈 계획이다.

“KISTI 같은 공공기관에서 시소러스를 구축하기 시작했다는 것은 국가적인 행운입니다. 일반 사기업에서 구축하는 시소러스만으로는 국가적 재산이 되기 힘들어요. 저희가 생각하는 가장 발전적인 모델은 출연 연구기관들이 각각의 분야에 관한 시소러스를 구축하고, 이러한 시소러스들을 모으고 활용하는 허브의 역할을 KISTI가 담당하는 것입니다. 그렇게만 된다면 국가적으로 엄청난 자원이 될 수 있을 겁니다. 미래에는 인간을 능가하는 체계적 지능을 갖춘 컴퓨터가 인간 대신 정보를 검색하는 시맨틱 웹이 보편화 될 텐데, 시소러스는 시맨틱 웹의 없어서는 안 될 중요한 기본 인프라입니다. 또 유비쿼터스 네트워크를 구성하는 언어 자원으로서도 시소러스는 없어서는 안 될 매우 중요한 인프라입니다. 이러한 분야에서 세계를 선도하려면 결국 보다 체계적이고 정확하고 방대한 시소러스를 구축하는 것이 무엇보다 중요한 선점 과제이고, KISTI와 같은 기관의 역할이 절대적으로 필요합니다.”

오름정보 이상현 사장은 이러한 측면에서 KISTI와의 협업은 오름정보에게 있어서 매우 소중한 기회라고 말했다. KISTI가 국가차원 시소러스 구축의 허브가 되고, 오름정보가 시스템구축의 중요한 파트너가 될 수 있다면 기업경영이나 수익 차원을 떠나서라도 더없이 보람을 느낄 것 같다는 것이다.

“KISTI를 통해 참 많은 것을 배웠습니다. 기업의 수익모델을 탄탄히 세우고, 기술적으로 자극을 받았던 것도 많았지만, 더욱 큰 도움이 된 것은 KISTI의 도전정신이었습니다. 정보의 수집과 가공, 서비스는 물론이고 연구자들에게 보다 효용성 높은 정보를 제공하기 위해 검색관리시스템과 시소러스까지 적극적으로 구축하려고 하는 도전정신이 많은 도움이 됐어요.”

끝으로 이상현 사장은 KISTI의 도전정신이 시소러스 분야에서 큰 성과를 낳을 것임에 틀림없으며, 설사 지금은 크게 부각되지 않더라도 앞으로 5년 뒤 10년 뒤에는 KISTI의 이러한 성과들이 국가 과학기술계의 더없이 소중한 자산이 될 것을 확신한다고 말했다.

그리고 시소러스 분야에 대한 오름정보의 확고한 의지도 피력했다. “완전함을 뜻하는 ‘오름’을 기업명으로 쓰는 만큼 모든 이용자들이 만족할 수 있는 완벽한 시소러스를 구축하기 위해 저의 오름정보 또한 최선을 다하겠습니다.” 

