

컴퓨터의 文字인식에 끌려드는 마력



박희선 <충북대 컴퓨터과학과 박사과정 2년>

문자인식에 대한 아주 미비한 지식만을 가진 채 시작된 나의 연구는 이제 나름대로의 연구영역을 가지고 조금씩 성장하고 있다. 돌이켜 생각해보면 그때의 난 얼마나 왜소했었던가!

대학원에 진학하면서 내가 받은 교육은 스파르타식 훈련 이상이었던 것 같다. 월요일부터 토요일까지 매일 아침 9시에 나와 빨라야 밤 10시에 연구실을 나섰다. 내가 지금까지도 계속해서 이렇게 생활 할 수 있는 것은 아마도 문자 인식이 내게 주는 어떠한 마력 때문일 것이다.

석사과정 1년차에 나는 연구의 시작을 어떻게 해야 하는가에 대해서 지도교수로부터 배웠다. 1년에 수십편씩 쏟아져 나오는 문자인식 관련 논문들을 수집하여 읽고 일주일에 적게는 한번, 많게는 두번꼴로 돌아오는 세미나에 논문을 발표하면서 많은 논문들을 내것으로 소화했다. 나의 지도교수는 내가 자신이 없는 부분을 슬쩍 넘어가려 할 때면 더욱 꼬치꼬치 캐물어 곤경에 빠뜨렸다. 그 당시에는 원망도 많이 했지만 이제는 그 분께 감사드린다. 이렇게 수십편의 논문을 발표하면서 가능성이 보이는 논문에 대해서는 곧바로 구현에 들어간다. 구현의 초기 단계는 논문을 충실히 따라가는 것이며 그 다음 내 문제에 적용한 후 개선하는 단계로 접어든다. 처음 나는 신선하고 기발한 아이디어를 짜내려고 몇번이나 시도한 적이 있었다. 그러나 다른 사람의 알고리즘을 그대로 구현함에 있어서도 저자의 아이디어들이 많은 부분 생략되어 있어서 가끔 한계에 부딪치기도 했다. 또한 그 방법에 대한 개선 실험에서 작은 아이디어조차 때로는 떠오르지 않을 때에는 심한 좌절감에 빠지기도 했다. 실험결과는 계속되는 연구세미나를 통해 거듭나게 되고 이후 이를 정리하여 논문을 작성했다. 학술회의에 먼저 발표하고 보다 가치있는 논문은 학술지에 투고했다. 이 과정에

서 논문의 문제점들이 심사위원들로부터 지적받기도 하는데, 한편으로는 다행스러우면서도 다른 한편으로는 또 이를 어떻게 풀어나가야 하는가로 한숨이 나오기도 했다.

내가 맡았던 최초의 문자인식프로젝트는 석사과정 1학년때 한국과학기술원 인공지능연구센터가 위탁한 '온라인 필기한자 인식 시스템의 개발'에 관한 연구였다. 3년전인 그 당시만 해도 온라인 문자인식에 관한 사람들의 인식은 극히 미약했고 국내에서 상품화를 염두에 두고 개발에 착수하기는 처음이었기 때문에 두려움이 나를 엄습했다. 특히, 국내의 온라인 필기한자인식에 관한 연구는 전무한 상태였던지라 나의 노동력을 필요로 하는 작업들이 연구초기에 물밀듯이 밀려들었다. 한자 데이터베이스 구축을 위한 획순자료집의 작성과 하루에도 몇백글자를 태블릿에 써야만 했던 일 등은 내가 했던 많은 작업들 가운데 가장 힘들었던 작업이었다.

석사과정 후반기부터 나는 오프라인 필기체 한글 인식으로 연구의 방향을 전환시켰다. 필기체 한글 인식은 문자수의 방대함과 문자간의 유사성이 심한 것 이외에도 인쇄체 문자에 비해 많은 변형이 따르는 등 어려움이 있었기 때문에 초창기 이 분야에 대한 연구 결과에서 큰 진전 없이 멈추어진 상태였다.

최근 나는 Markov Random Field 모델에 강한 호기심을 가지고 있다. 아직은 문자인식 분야에서 생소한 모델이지만 충분히 가능성이 있을 것이라는 믿음을 가지고 오늘도 이와 관련 논문들을 읽고 기본 이론들을 이해해 가면서 문자 인식 분야에 응용하고 있다.

문자인식의 문제 해결을 위해 내가 얼마만큼 보탬이 될 수 있을지에 대해서는 미지수이나 분명한 것은 젊은 시절 도전해 볼 만한 가치있는 연구분야라는 사실이다. 