

재미있는 우편상식+

순로구분기가 뜬다

박 종 흥*

인터넷, GPS, WiFi와 같은 용어는 더 이상 전문가들의 전유물이 아니다. 10년 전 만해도 전문가 그룹, 그 전에는 군과 같은 특정 집단에서만 알고 있는 고도의 기술적 용어였지만 휴대폰, 인터넷, 내비게이션과 같은 첨단기기들이 보편화되면서 평이한 용어가 되어 버렸다. 이와 같이 한 때는 '기술'에 해당되는 전문지식이 시간이 지나면 상식이 되곤 한다.

우정정보지의 특성상 '기술'이라는 단어가 들어간 제목을 쓴다면 독자들은 아예 들춰보지도 않을 것 같다. 그래서 본 칼럼을 기술보다는 편안한 그래도 상식보다는 조금 전문성이 있는 내용을 다루기 위해 '상식+'란 제목으로 만들어 보았다. 우편의 전반적인 업무와 전문 지식이 필요한 독자층에게 많은 도움이 되길 바란다.

얼마 전 순로구분기 운영 매뉴얼이 배포되었다. 매뉴얼이란 것이 양산되는 제품에 따라오는 사용 지침서란 점에서 매뉴얼이 작성되었다는 것은 이 기계가 그 만큼 많이 보급되고 있다는 의미다. 또한 제품이 널리 사용되는 만큼 올바르게 운영해야 제 성능을 발휘할 수 있으므로 이제 순로구분기는 우체국 근무자라면 반드시 알고 있어야 하고 기계 운영자라면 속속들이 기계 내부를 파악하고 있어야 하는 대상이 되었다.

순로구분기는 영어로 Carrier Sequence Sorter다. 집배원(Carrier)의 배달 순서(Sequence)를 자동으로 정렬해주는 기계(Sorter)로 선진국들은 이미 십여 년 전부터 사용하고 있다. 우리나라는 2005년부터 시흥우체국과 노원우체국의 시범 사업을 필두로 보급되기 시작했고 관련

* 한국전자통신연구원(ETRI) 우정물류기술연구부장, jpark@etri.re.kr

기술은 2001년부터 약 5년간에 걸친 대대적인 사업을 통해 개발되었다. 올 해 120여 대가 운영되고 있고 앞으로 통상우편 물량 변동에 맞추어 수백 대는 더 보급될 것으로 보인다. 기계 수명과 물량 면에서 하나의 산업이 당분간 선순환될 것으로 보인다.

순로구분기가 탄생되는 과정을 보면 그리 순탄하지는 않았다. 10년 전 쯤 외국에서 운영되는 기계를 보고 과연 저런 엄청난 기계, 초당 7내지 8통의 우편물이 “타르르르...” 소리를 내면서 지나가면 그 순간에 주소를 읽고 수백 개의 구분칸에 정확히 우편물을 떨어뜨리는 그런 기계를 과연 우리 실력으로 개발해 낼 수 있을까? 당시 많은 관계자들이 이런 의구심 가졌던 것은 당연한 일일지 모른다. 그래도 한글을 인식하는 기술은 한국에서 개발되어야 하고 가장 잘 할 수 있을 것이란 믿음과 도전정신으로 일은 벌어졌다. 5년에 무려 90여억 원의 개발비가 들어갔고 년 10여명의 연구원들이 투입되었다. 개발 초기 2, 3년은 결과물을 보여줄 수 없어서 “돈 먹는 사업”이니, “차라리 그 돈으로 기계를 도입하는 것이 효과적이다”라는 날카롭고 험한 질책도 있었다. 그럼에도 불구하고 우정사업본부의 확고하고 꾸준한 지원과 ETRI 연구원들의 열정으로 지금의 결실을 맺을 수 있었다. 과급효과를 제외한다고 해도 순로구분기 도입에 거의 1,000억 원의 비용이 절감되었으니 10배가 넘는 장사를 한 셈이다.

그러면 도대체 순로구분기에는 어떤 기술이 들어가 있기에 그 많은 시간과 돈을 들여서 개발해야 했을까? 순로구분기를 한 꺼풀 벗겨보기로 하자.

순로구분기는 제작사별로 하드웨어적인 차이는 있으나 전체 구성은 비슷하다. 구분기를 제어하고 진단하는 관리시스템, 우편물을 이동하고 기계적으로 구분해서 적재하는 기계적 본체, 주소 정보를 기준으로 순로를 정의해주는 구분계획 관리시스템, 우편 영상을 판단하고 미처리 영상을 오프라인으로 보내주는 우편영상 정보관리 시스템, 우편 영상으로부터 한글주소를 추출하고 정확한 주소를 읽어내는 한글주소인식 클라이언트, 주소 정보 DB를 관리해 주는 주소정보 관리시스템이 그 핵심 구성 요소이다.

모든 구성품이 난이도가 있는 기술이지만 가장 어렵고 중요한 것은 한글주소인식 클라이언트에 들어가 있는 한글 인식과 주소해석 기술이다. 집중국의 서장구분기는 지역을 대표하는 우편번호만 읽고 도착지별로 구분해주면 되지만 집배국의 순로구분기는 사람과 똑같이 주소를 읽어

야 하기 때문에 이 기술이 차지하는 비중이 절대적이다. 또한 우편물 주소를 그냥 읽기만 하면 되는 것이 아니고 사람보다 더 빠르게 읽고, 주소DB 매칭으로 틀린 글자까지 고쳐주어야 한다. 더구나 이러한 일련의 작업을 1/10초 이내에 끝내야 하는 녹녹치 않은 기술이다.

주소를 읽는 과정을 설명하기 위해 잠시 '인식'이 무엇인지에 대해 생각해보자. 사전적인 의미의 인식이란, '사물을 분별하고 판단하여 아는 일'로 정의 되어 있다. 곧 어떤 관찰 대상을 양자화된 하나의 개념에 연결하는 것, 또는 그 과정이라 볼 수 있다. 그럼 쉬운 예로 우리가 사람 이름을 인식하는 과정과 비교해 보자.

갑순이가 집에 돌아왔더니, 기억력이 좋지 않은 할머니께서 갑돌인가, 갑도인가라고 하는 사람에게서 전화가 왔었다는 말씀을 전해 주신다. '갑돌'이라는 사람의 이름을 '갑돌' 또는 '갑도'라는 이름으로 잘못 입력받는 상황이다. 갑순이는 금세 '갑돌'이가 전화했음을 알 수 있다. 어째서 그럴까? 이러한 상황에서 우선, 우리의 뇌는 알고 있는 사람들의 이름을 주욱 훑으면서, 입력 받은 이름과 매치시켜 보는 과정을 자연스럽게 진행한다. 그런 다음 저장된 사람 이름 목록에서 '갑돌'이라는 사람이, 입력받은 '갑돌' 또는 '갑도'와 가장 유사한 이름임을 찾아내고, 갑돌이라는 사람을 생각해 내게 된다. 이 과정이 바로 인식이다.

[그림 1] 언어의 인식

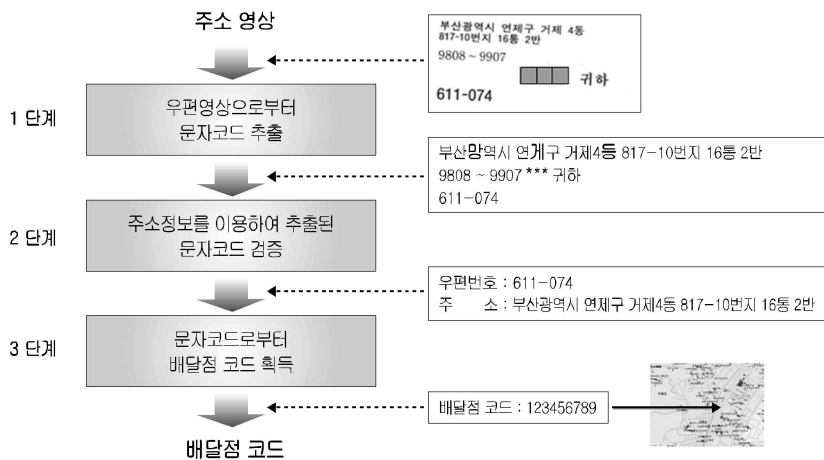


주소 인식도 마찬가지다. 우편물 처리 과정에서 주소 인식은 주소가 기재된 영상에서 주소 문자들을 추출해 내고, 이것들과 가지고 있는 데이터들을 비교해 해당 주소지를 찾아내거나 가장 그럴싸한 주소지로 연결시킨다. 물론 좀 전의 예는 음성 인식에 가깝고, 우편물 처리 과정에서는 영상 인식이 진행되지만, 입력되는 형태가 소리나 영상이나의 차이만 있을 뿐 입력된 데이터에서 어떤 사실이나 사물을 인식하는 과정은 비슷하다.

순로구분기에서 동작하는 주소 인식은 크게 세 단계로 나눌 수 있다. 1단계는 그야말로 맞거나 틀리거나 읽기만 하는 단계다. 검은 색은 글자고 흰색은 바탕이니 검정색 기준으로 주소가 기입된 영상에서 해당 주소 문자 코드만을 알아내는 것이다. 곧 주소 정보를 이용하지 않고, 인식해야 할 범위에 있는 모든 글자를 영상에서 뽑아 낸 뒤 나열한다. 예를 들어 '부산광역시 연제구 거제 4동 817-10번지 16통 2반 9808~9907 * * * 귀하 611-074'라는 모든 문자 추출은 단순한 문자 인식에 지나지 않는다.

다음으로 2단계는 주소가 기입된 영상에서 주소 문자열을 뽑는 과정이다. 여기에서 가지고 있는 주소 정보로 잘못 인식된 문자를 수정하거나 주소 인식 결과를 문자열로 처리, 추출된 문자열 가운데 주소와 관련 있는 문자열인 '611-074'와 '부산광역시 연제구 거제 4동 817-10번지 16통 2반'만을 결과로 돌려주는 것이다.

[그림 2] 주소 인식의 단계

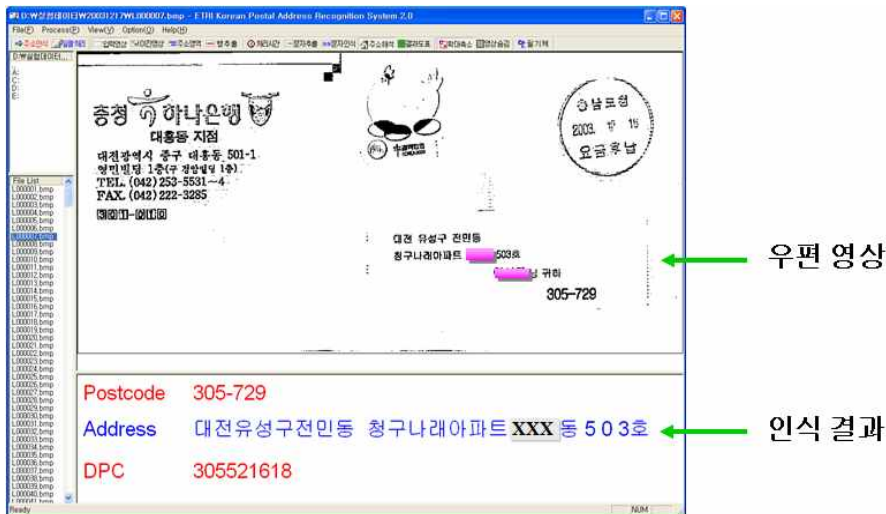


마지막인 3단계는 주소가 기입된 영상에서 주소지를 알아내기까지의 과정이다. 입력받은 영상에서 주소 정보를 인식해 주소의 ID를 알 수 있고, 실제 위치 정보나 주소지 정보를 뽑을 수 있는 정도의 수준이다.

이와 같이 순로구분기에 탑재된 한글 주소 인식 시스템은 단순히 주소에 적힌 한글을 읽는 수준을 넘어, 실제 주소지 정보를 얻을 수 있는 3단계 모두를 포함한다.

가끔 이 기술을 설명하다 보면 한글인식은 이미 다 실용화되어 있다고 주장하는 사람을 만나게 된다. 명함 인식 시스템이나 차량 번호 인식 기술 등과 동일선상에서 비교해서 이러한 주장을 하게 되는 것이다. 하지만 이들은 모두 앞에서 언급한 1단계 인식에 지나지 않아 잘못 읽혀도 읽힌 그대로 처리하는 수준이지 주소를 보정하거나 ID까지 찾아주는 시스템은 아니다. 따라서 순로구분기에 들어가는 기술은 단순 글자 인식과는 다르게 지능화되어 있고, 운영하는 과정에서 점점 더 성능이 좋아지는 학습 효과까지 가지고 있어서 기존의 인식 기술과는 차원이 다르다.

(그림 3) 한글 주소 인식 시스템의 인식 결과



순로구분기가 우편물의 주소를 읽어서 구분해 주는 성능은 한 시간에 3만 통이 넘고 구분율은 85%가 넘는다. 순로구분 성능은 시간당 약 8천 통이다. 구분 성능과 순로 구분 성능이 다른

이유는 우편물을 순서대로 정렬하기 위해서는 주소를 읽은 후 몇 단계를 더 거쳐야 하기 때문이다. 수학을 좋아하는 사람이라면 쉽게 이해하겠지만 3자리 수 카드 10장을 10개의 통을 이용해 순서대로 정렬하려면 3번의 반복적인 구분 작업이 필요하게 된다. 이를 기수 정렬방식(이 방식에 대한 상세한 내용은 순로구분기 매뉴얼 23쪽 참조)이라고 하는데 마찬가지로 주소를 읽은 후 배달 순서까지 정렬해 줄려면 구분 칸의 개수를 고려해서 세 번을 돌려야 한다. 따라서 순로구분 성능은 구분 성능의 1/3 정도가 되는데 부대시간을 감안해서 계산해 보면 시간당 8천 통이 된다.

순로구분기 이후 한글 주소 인식 기술은 서장구분기에도 탑재되고 있다. 서장구분기가 단순 우편번호 인식에서 탈피해 주소까지 인식하게 함으로서 판독의 정확성을 높이고 필요하면 소규모 집배국의 순로를 구분하는 기능까지 보유할 수 있게 하는 것이다. 나아가 앞으로는 수신자명을 기준으로 전송이나 반송 우편물을 사전에 찾아내어 번거로운 수고와 낭비를 방지하는 등 주소 기반의 우편 체계가 정착될 수 있게 된다. 또한 이 기능을 서장구분기의 기본 요구사항으로 정의함으로써 국내 기업의 경쟁력이 높아지는 효과를 볼 수 있다.

이로서 우리나라는 순로구분기, 서장구분기, 등기통상구분기와 같은 통상우편물용 자동구분기를 100% 국내 기술로 생산할 수 있게 되었다. 우편관계자 뿐만 아니라 국내 관련 산업계로선 큰 쾌거가 아닐 수 없다. 남은 일은 수출이다. 단지 상대 국가가 알파벳이 아닌 문자를 사용하는 경우 해당 국가의 문자 인식 기술과 주소 데이터베이스 구축이 필요한데 이를 위해서는 우리의 원천 기술을 전수하는 방법도 있다. 이 경우 외산과 비교해 볼 때 성능은 동등하고 가격은 저렴하니 국제 경쟁력이 상당히 높아진다. 언제 누가 어떻게 물꼬를 트느냐가 관건이다.

참 고 문 헌

- 박종홍, 김인수, 엄보운, 「흰히 보이는 우편기술」, 전자신문사, 2009. 12.
우정사업본부, 『순로구분기 운영매뉴얼』, 2010. 7.