

# 콜센터를 위한 전자메일 자동응답 및 관리 시스템 설계 및 구현

김정재<sup>0</sup> 권오식<sup>†</sup> 이호동<sup>‡</sup> 노영균<sup>‡</sup> 박진영<sup>‡</sup> 박종철<sup>‡</sup>  
한국과학기술원 전산학과<sup>†</sup> MPC<sup>‡</sup>

{jjkim,oskwon,hdlee,park}@nlp.kaist.ac.kr<sup>†</sup> {yknoh,jypark}@callcenter.co.kr<sup>‡</sup>

## Design and Implementation of E-mail Response Management System for Call Center

Jung-Jae Kim<sup>0</sup> Oh-Shik Kwon Ho-Dong Lee Young-Kyun Noh  
Jin-Young Park Jong C. Park  
Dept. of Computer Science, KAIST MPC

### 요 약

본 논문에서는 콜센터를 위하여 설계 및 구현된 전자메일 자동응답 및 관리 시스템 중에서 서버 시스템에 해당하는 부분을 기술하였다. 본 연구에서는 도메인에 특성화된 표현 형식 개발을 개발하여 보다 효율적인 3단계 매칭방법을 가진 자동응답기, 학습에 기반한 도메인 비의존적인 자동분류기 및 적용방법의 재배열이 가능한 담당자 분배기를 구현하였다.

### 1. 서론

최근 인터넷이 급격하게 대중화됨에 따라 전자메일이 전화와 같은 일반적인 의사전달 수단으로 자리잡고 있다. 이에 전자메일이 가지는 상대적인 편의성이 추가되어 그 이용이 앞으로 폭증할 것으로 예상된다. 그 예로 해당 기업의 고객이 주로 전화 상담 서비스를 통하여 그 기업의 제품이나 서비스에 대해 문의를 하는 기존 양상이 빠른 속도로 전자메일을 이용한 문의로 대체되어 가고 있는 것을 들 수 있다. 이렇게 전자메일을 이용한 문의의 규모가 커지게 되면 지금과 같이 상담원들이 고객의 전자메일에 일일이 응답을 하는 방식으로는 접수에서 답변까지 지체되는 시간이 점차 증가하게 되며 처리가 되지 않고 방치되는 전자메일의 수효도 지속적으로 증가하게 되며 이는 해당 기업의 효과적인 고객관리에 바람직하지 않은 결과를 초래하게 된다. 본 논문에서는 콜센터에서 사용할 수 있는 전자메일 자동응답 및 관리 시스템의 개발에 대하여 논의한다.

폭증하는 전자메일을 처리하기 위하여 고려할 수 있는 방안으로 충분한 상담원을 확보하여 전자메일을 직

접 관리하는 방안과 자동/반자동화된 전자메일 분류 및 응답 시스템을 이용하는 방안이 있다. 상담원을 이용하는 방안은 보다 친숙한 서비스를 고객에게 제공할 수 있다는 장점이 있는 반면 폭증하는 전자메일의 양에 비례하는 규모의 상담원을 미리 확보해야 하고 상담원의 초기 교육 및 질적인 서비스의 유지에 많은 자원을 투자해야 한다는 단점이 있다. 현재 전세계의 많은 기업들이 자동화된 전자메일 분류 및 응답 시스템을 이용하는 방안에 보다 높은 기대와 함께 전략적인 타당성을 부여하고 있다.

본 연구에서는 콜센터를 위한 전자메일 자동응답 및 관리시스템을 개발할 때 반드시 요구되는 사항으로써 다양한 도메인에 쉽게 적용할 수 있는 도메인 비의존적 특성과 특정한 형태의 메일에 대한 정확한 답변을 요구하는 도메인 의존적 특성을 함께 고려하였다.

### 2. 전자메일 자동관리 시스템 구조

본 연구에서 기초로 하는 전자메일 자동관리 시스템의 전체적인 구조는 그림 1과 같다.

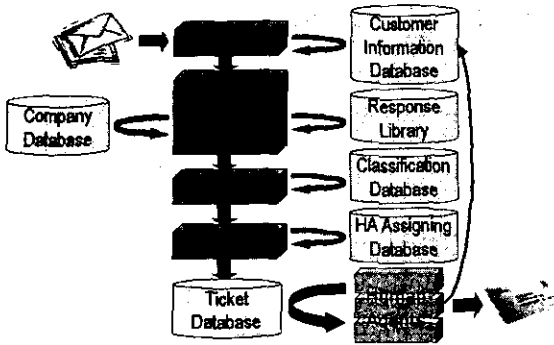


그림 1 ERMS 전체 시스템 구조

고객이 보낸 메일을 전자메일 수신기, 자동응답기, 자동분류기, 자동분배기를 통하여 필요한 내용을 자동 추출하여 데이터베이스에 저장하면 메일담당자는 자신이 담당하는 클래스의 메일을 읽어들이 필요한 답변내용을 작성한 다음 고객에게 답장한다.

이와 같은 구조에서 자동응답기는 고객의 문의메일 중에서 특정한 형태 또는 의미를 지닌 메일들을 찾아내어서 그에 맞는 답변을 자동추출하게 된다. 자동분류기는 자동응답기가 처리하지 못하는 메일들에 대하여 가장 적절한 클래스를 선택하여 줌으로써 특성화된 메일담당자에게 메일을 분배할 수 있도록 한다. 이때 자동분류기의 경우 모든 메일에 대하여 적절한 클래스를 지정해 주어야 하는 반면 자동응답기의 경우 되도록 자동응답의 결과가 정확해야 하는 제약을 받는다.

### 3. 도메인 비의존적 자동분류기

자동분류기는 메일 분류 항목 생성, 메일의 분류 항목으로의 할당, 학습, 그리고 실시간으로 이루어지는 메일 분류 순으로 진행된다.

자동분류기는 주로 명사를 키워드로 하여 명사가 분류 항목마다 어떻게 분류되어 있는가를 파악하여 분류하고 학습된 키워드들은 각 클래스에 대한 점수를 원소로 하는 벡터를 가진다. 새로운 메일이 들어오면 메일 내용에 포함된 학습된 키워드들의 벡터 합으로 나

타내어지는 벡터로, 메일을 표현할 수 있으며, 이 벡터의 방향으로 메일의 클래스를 선택한다.

키워드가 어떤 집단에 크게 기여한다는 것은 그 키워드가 그 집단에 집중적으로 나오고, 다른 집단에는 상대적으로 적게 나오는 것을 의미한다. 본 연구에서는 이를 위하여 TF (The Within-Document Word Frequency)와 IDF (The Inverse Document Frequency)를 이용하였다.

하나 이상의 클래스를 포함하는 클래스 집단과 클래스 간의 관계는 또 다른 형태의 행렬구조로 표현된다. 이는 종속관계, 독립관계를 표현하는 행렬로서 이진 데이터를 사용한다. 즉, 클래스 집단에 포함된 클래스는 1, 포함되지 않은 클래스는 0으로 나타내어진다. 이를 바탕으로 임의의 클래스 계층구조는 클래스 집단을 클래스 항목으로 관계지을 수 있다.

이상에서 얻어진 키워드와 클래스 집단간의 관계 행렬을 클래스 계층구조 행렬에 매핑시키는 변형행렬을 구함으로써 키워드들의 클래스 집단에 대한 기여도를 구할 수 있다.

실시간 클래스 분류는 클래스 별로 자주 나타나는 키워드가 편중되는 경향을 보이며, 메일에 쓰이는 단어 수가 제한될 것이라고 가정한다. 본 연구에서는 이와 같은 자동분류기가 높은 성공률을 얻기 위해서 분류기 기준의 중복 불가, 분류기준들 사이의 상하 종속 관계의 재정립, 분류 기준의 깊이 제한 등을 고려하여 분류 항목들을 정의한다.

### 4. 도메인에 특성화된 자동응답기

본 연구에서 선택한 전자메일 데이터에 대한 분석결과 자동응답시 변수가 되는 단어나 구의 경우, 예를 들어 제품명과 같이 일련의 정해진 키워드 또는 구의 리스트, 전화번호나 전자메일주소와 같이 숫자나 영자모의 일정한 형태의 나열 또는 주소나 금액과 같은 일정한 형태의 패턴으로 표현될 수 있었다. 본 연구에서는 이러한 형태의 패턴들을 간결하게 표현하기 위하여 숫자의 연속, 영자모의 연속 등의 도메인 표현에 적합한 변수들을 정의하여 패턴들을 표현하였다.

그리고 고객이 작성한 메일들은 맞춤법 오류, 띄어쓰기 오류 등의 치명적인 오류를 많이 담고 있으며 메일에서 많이 발견되는 극도의 축약형태, 문장구분자의 부재, 일반문장에서 사용하지 않는 기호, 완전한 문장 형태가 아닌 표의 내용을 텍스트로 표현한 것 등의 특성을 가지고 있어서 전자메일의 처리를 위해서는 문장 전체에 대한 완전과정보다는 부분 패턴매칭이 보다 적합한 것으로 판단된다.

본 연구에서는 이러한 메일의 특성을 감안하여 다음과 같은 3단계 매칭방법을 채택하였다.

A. Unit Matching : 자동응답시 중요한 변수가 될 수 있는 단어나 구 등을 일정한 패턴 또는 리스트, 즉 unit으로 표현하여 매칭한다.

#### B. Signature Matching

같은 의미의 여러 다른 형태의 표현을 하나의 의미로 묶어주고 멀리 떨어져 있는 의미들을 따로 분석한 다음 하나로 합쳐주기 위하여 같은 의미의 여러 표현들을 한 단위, 즉 signature로 묶어서 표현하였다. 이러한 signature를 매칭할 때 전체 문장을 완전히 과상하는 것보다는 특정한 표현이 일정한 형태로 배열되었는지를 우선적으로 검사한다. 이때 일정하지 않는 줄구분대상으로 견고한 매칭을 하기 위하여 한 줄을 뛰어넘어 매칭하는 operator를 도입하였다.

#### C. Case Matching

일반적으로 하나의 자동응답에 하나의 case가 존재한다. Case는 여러 개의 signature가 하나로 합쳐져서 이루어졌으며 signature에서 정의된 변수들을 이용하여 case의 답변내용을 기술한다. 또한 답장을 보내기 전에 먼저 처리해야 하는 일들도 case에 기술된다.

이와 같은 3단계 매칭방법 이외에 또 다른 중요한 기능은 자동응답시의 DB와의 연동이다. 예를 들어 고객이 회사에 특정 제품에 대한 재고여부를 문의하였을 경우 그 제품의 재고가 남아있는지의 여부에 따라 답장내용은 달라져야 한다. 또한 고객이 주문한 상품들의 배송상황을 문의하였을 경우 답장은 각 상품들의 배송상황 정보를 포함하여야 하는데 이 정보는 DB와의 연동에서만 얻어질 수 있다.

### 5. 적용방법의 재배열이 가능한 자동분배기

본 논문에서는 각 메일에 대한 메일담당자를 다음의 4가지 분배방법을 이용하여 분배하도록 하였다.

- 1) N명의 메일담당자에게 메일 하나씩 분배하여 메일담당자마다의 할당 메일 수가 같게 한다.
- 2) 메일의 클래스를 담당하는 메일담당자에게 메일을 분배한다.
- 3) 메일을 보낸 특정 고객을 담당하는 메일담당자에게 메일을 분배한다.
- 4) 메일을 보낸 고객이 가장 최근에 보낸 메일을 담당한 메일담당자에게 메일을 분배한다.

이와 같은 적용방법이 도메인의 특성에 따라 재배열이 필요하므로 본 연구에서는 이들간의 재배열 및 적용여부를 관리자가 선택할 수 있도록 하였다.

### 6. 결론

본 논문에서는 콜센터를 위하여 설계 및 구현된 전자메일 자동응답 및 관리 시스템 중에서 서버 시스템에 해당하는 부분을 기술하였다. 본 연구에서는 도메인에 특성화된 표현 형식 개발을 개발하여 보다 효율적인 3단계 매칭방법을 가진 자동응답기, 학습에 기반한 도메인 비의존적인 자동분류기 및 적용방법의 재배열이 가능한 담당자 분배기를 구현하였다.

앞으로의 시스템 개선 방향으로 자동분류기 기능 세 부화, DB 연동 기능 개선, 상이한 전자메일 코퍼스에 대한 시스템 적용 및 성능 개선방안 연구 및 사용자 중심 자동응답 기능 제공 등이 있다.

### 7. 참고문헌

- [1] D. Fluss, D. Hope-Ross, T. Berg, E-Mail Response Management Systems Market Analysis, GartnerGroup Strategic Analysis Report, 1999.
- [2] Yiming Yang and Christopher G. Chute, An Example-Based Mapping Method for Text Categorization and Retrieval, ACM Transactions on Information Systems, 12(3):252-277, 1994.
- [3] Ellen Riloff and Wendy Lehnert, Information Extraction as a Basis for High-Precision Text Classification, ACM Transactions on Information Systems, Vol. 12, No. 3, pp. 296-333, 1994.