의미추론규칙을 이용한 온톨로지 기반의 스팸방지 시스템

허정환[O]* 정진우* 주영도** 이동호*
한양대학교 컴퓨터공학과*, 강남대학교 컴퓨터미디어공학부**
{hyugar[O], jwjeong, dhlee72}@hanyang.ac.kr*, ydjoo@kangnam.ac.kr**

Ontology-based Anti-Spam System using Semantic Inference Rules

Chung-Hwan Heu[O]* Jin-Woo Jeong* Joo, Young Do, ** Dong-Ho Lee*
Dept. of Computer Science & Engineering, Hanyang University*
Division of Computer and Media Engineering, Kangnam University**

요 약

전자우편(email)은 인터넷의 급격한 보급으로 인하여 사용자들이 많이 사용하게 된 통신 메커니즘이다. 그러나 이러한 전자우편의 대중성을 상업적인 목적으로 이용한 스팸메일의 출현으로, 사용자들은 정신적 피해, 업무 방해, 메일서버의 트래픽 과부화로 인한 유지보수 비용 증가와 같은 문제점들을 접하게 되었다. 특히, 최근에는 광고성 이미지들을 첨부하는 등의 새로운 기법이 적용된 스팸메일의 발생으로 기존의 텍스트 기반의 스팸메일 필터링 기법들이 무의미하게 되었으며, 따라서 그로 인한 피해가 증가하는 추세이다. 이러한 이미지 기반의 스팸메일들의 필터링을 위하여 Support Vector Machine과 같은 기계학습 기법을 이용한 기법들이 제안되고 있으나, 여전히 그 성능은 만족스럽지 못하다.

본 논문은 전자우편으로부터 텍스트 및 시각적 의미를 분석하여 전자우편 온톨로지에 기술하고 스팸메일 판단을 위한 의미추론규칙을 적용함으로써 광고성 이미지가 첨부되어 있는 스팸메일을 효과적으로 필터링 하기 위한 시스템을 제안한다.

1. 서 론

전자우편은 인터넷의 급격한 보급으로 인하여 사용자들이 많이 사용하게 된 통신 메커니즘이다. 통계에 따르면 2006 년을 기준으로 전 세계에 약 600 억 통의 전자우편이 발송되고 있다고 한다[1].

그러나 이러한 전자우편의 대중성을 상업적인 목적으로 이용한 전자우편인 스팸메일이 발생되었다. 스팸메일은 본인이 원치 않음에도 불구하고 일방적으로 전송되는 영리목적의 광고성 메일로, 최근 조사에 따르면 전체 전자우편 트래픽 중 60%가 스팸메일이라고 한다[1].

스팸메일로 인한 피해는 크게 사용자 측면과 관리자 측면으로 나눌 수 있다. 사용자 측면에서 보면 첫째, 공격적이고 자극적인 광고물들로 이루어진 스팸메일로 인하여 사용자들은 정신적인 피해를 입게 되며, 둘째, 반복적으로 대량의 스팸메일이 발신되기 때문에 필요한 정보의 수신을 방해 받게 된다. 셋째, 빈번하게 발생하는 스팸메일을 삭제하는 작업은 업무에 많은 지장을 초래 하게 된다. 관리자 측면에서 보면 첫째, 스팸메일은 반복적으로 대량전송이 되기 때문에, 메일서버의 트래픽 과부하를 초래하는 문제가 발생되게 된다. 둘째, 필요 없는 전자우편들을 대량으로 저장하기 때문에 이를 관리하는 메일서버의 저장 공간을 추가적으로 구입해야 하는 문제가 발생된다. 통계에 따르면 메일서버 운영 업체에서는 저장 공간의 추가 구입을 하기 위하여 매년 약 100 억 달러의 비용을 추가한다고 한다[1].

스팸메일로 인한 피해를 줄이기 위하여, 전자우편을 분류하여 필터링 하기 위한 다양한 기법들이 제안되었다. 초기 스팸메일은 제목과 본문에 상업적인 텍스트들을 삽입하는 방식이 대부분을 차지하였다. 연구자들은 전자우편의 제목과 본문에 있는 텍스트들의 빈도를 분석하고 SVM 등과 같은 기계학습을 이용하는 필터링 기법들을 제안하였다. 그러나 스팸메일 발신자들은 이러한 필터링 기법을 피하기 위하여 스팸메일에 광고성 이미지를 첨부하는 새로운 기법을 적용한 스팸메일들을 발신하기 시작하였다. 최근에는 이러한 새로운 기법들이 적용된 스팸메일을 효과적으로 필터링 하기 위한 다양한 연구들이 수행 중에 있다.

본 논문은 전자우편으로부터 텍스트 및 시각적 의미를 분석하여 전자우편 온톨로지에 기술하고 스팸메일 판단을 위한 의미추론규칙을 적용함으로써 광고성 이미지가 첨부되어 있는 변칙 스팸메일을 효과적으로 필터링 하기 위한 시스템인 OASIS(Ontology-based Anti-

Spam System using Semantic Inference Rules)를 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 기존 연구들에 대하여 알아보고, 3장에서는 본 연구에서 제안하는 방법에 대하여 상세히 설명한다. 마지막으로 4장에서는 결론 및 향후 연구 계획에 대하여 기술 한다.

2. 관련연구

초기 텍스트 기반의 스팸메일을 필터링 하기 위하여 연구자들은 전자우편 각 단어들의 빈도수를 계산하고, SVM, Naïve Bayesian 등과 같은 기계학습 분류자를 학습시켜 필터링 모델을 생성시키는 기법을 제안하였다[2.3]. 그러나 스팸메일 발신자들은 텍스트 기반의 필터링 기법을 피하기 위하여 전자우편에 광고성 이미지들을 첨부하는 변종 스팸메일들을 발신하기 시작하였고, 이러한 변종 스팸메일들을 효과적으로 필터링 하기 위한 다양한 기법들이 제안되었다.

Giorgio Fumera 의 연구[1]에서는 첨부 된 이미지의 텍스트들을 광학문자판독기를 사용하여 추출하고, 전자우편의 제목과 본문에 있는 텍스트들을 추가하여 빈도를 분석한다. 이 연구에서 저자들은 스팸메일 학습을 위한 트레이닝 집합으로부터 각 텍스트들의 빈도를 분석하고, 이를 이용하여 SVM 분류자를 학습시키고, 생성된 모델을 이용하여 스팸메일을 필터링 하는 기법을 제안하였다. 그러나 이 기법은 이미지가 첨부된 메일임에도 불구하고 필터링 과정에서 오직 전자우편에 존재하는 텍스트들의 빈도만을 이용하였다. 또한 제목과 본문에는 텍스트가 없고, 이미지만 첨부된 전자우편은 텍스트의 빈도수 계산만으로는 정확한 필터링 결과를 기대하기 어렵다. 그러므로 보다 정확한 스팸메일 필터링을 위하여, 스팸메일에 첨부된 광고성 이미지들만의 시각적인 패턴들을 파악하고, 그 특징들을 분석할 필요가 있다.

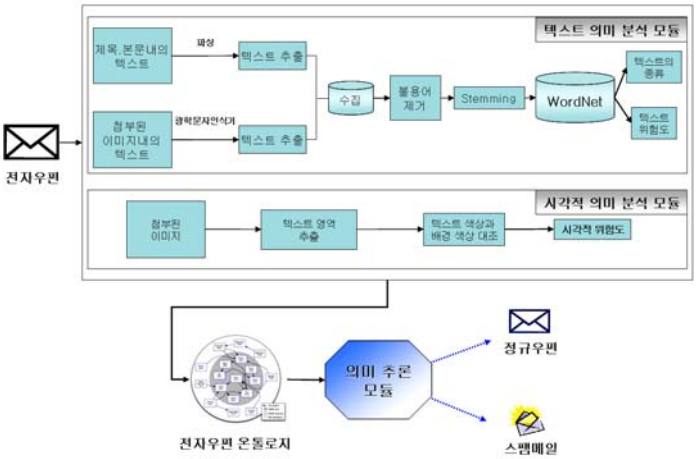
Hrishikesh B. Aradhye 의 연구[4]에서는 첨부된 이미지 안에 있는 텍스트의 범위, 색상의 순도, 색상 분포도의 정보들을 분석하고, SVM 을 통한 기계학습을 이용하여 스팸메일을 필터링 하는 기법을 제안을 하였다. 또한 Ngo Phuong Nhung 의 연구[5]에서는 윤곽선의 방향과 상호관계를 분석하고, 분석된 결과로 SVM 기계학습 분류자를 학습시켜 스팸메일을 필터링 하는 기법을 제안하였다.

[1,2,3,4,5]의 연구와 같이, SVM, Naïve Bayesian 등과 같은 기계학습 분류자들을 사용한 통계 기반의 스팸메일 필터링기법들은 분류결과가 비교적 정확하고 우수하다는 장점이 있지만, 필터링에 사용된 모델을 알 수 없고, 모델에 대한 분석이나 관리가 용이하지 않다는 단점이 존재한다. 또한, 기계학습으로 생성된 모델은 사용자의 취향을 고려하지 않고 필터링을 수행하기 때문에, 사용자의 취향에 따라 스팸메일의 종류가 변경되거나 추가 될 때 마다 높은 비용의 트레이닝 과정을 매번 수행하여야 한다. 즉, 기계 학습기반의 기법들은 사용자 취향에 따른 스팸메일 필터링에는 취약하다는 한계가 있다.

본 논문은 앞서 언급된 기존 필터링 기법들의 단점을 보완하기 위하여 전자우편으로부터 텍스트 및 시각적 의미를 분석하여 전자우편 온톨로지에 기술하고 스팸메일 판단을 위한 의미추론규칙을 적용함으로써 광고성 이미지가 첨부되어 있는 스팸메일을 효과적으로 필터링 하기 위한 시스템인 OASIS(Ontology-based Anti-Spam System using Semantic Inference Rules)를 제안한다.

3. 제안하는 시스템

3.1. 시스템구조

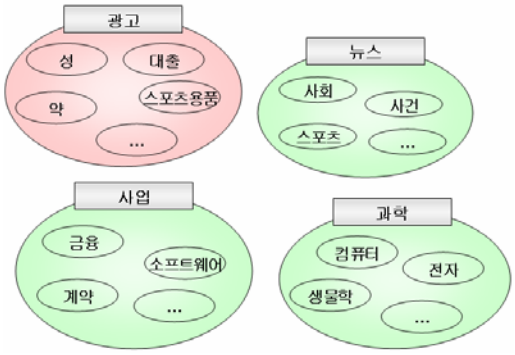


(그림 1) OASIS 시스템

제안하는 시스템의 구조는 (그림 1)과 같다. 시스템은 텍스트 의미 분석 모듈과 시각적 의미 분석 모듈, 분석된 정보들을 전자우편 온톨로지에 기술하여 의미추론규칙에 적용시키는 의미추론 모듈의 세가지로 모듈로 구성 된다. 텍스트 의미분석 모듈은 전자우편의 제목, 본문 그리고 첨부된 이미지의 텍스트들을 추출하여 전자우편이 속하게 될 종류와, 해당 전자우편의 스팸메일 가능성(위험도)를 분석한다. 마찬가지로, 시각적 의미분석 모듈은 첨부된 이미지의 색상의 대조도를 분석하여 해당 전자우편의 시각적 위험도를 분석한다. 마지막으로, 텍스트 의미 분석 모듈과 시각적 의미 분석 모듈을 통한 전자우편의 종류와 위험도를 전자우편 온톨로지에 기술하며, SWRL 기반의 의미추론규칙을 적용하여 최종적인 스팸메일의 여부를 판단하게 된다.

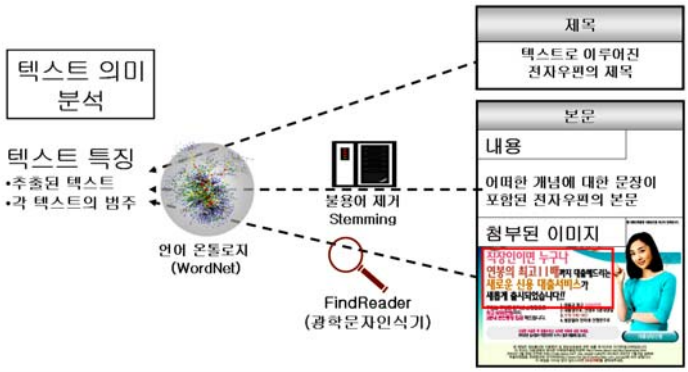
3.2. 텍스트 의미 분석

텍스트 의미분석 단계에서는 스팸메일을 판단하기 위하여 전자우편의 제목과 본문의 텍스트를 파싱 및 수집하고, 첨부된 이미지 내의 텍스트들을 추출한다. 추출된 텍스트들로부터 해당 전자우편이 속하게 될 종류와 해당 전자우편의 텍스트 위험도를 분석한다.



(그림 2) 전자우편의 범주와 종류

(그림 2)는 전자우편이 속하는 종류와 범주를 표현하고 있다. 전자우편의 범주는 전자우편이 가지고 있는 성질을 의미하며, 종류는 전자우편이 가지고 있는 내용을 의미한다. 예를 들어, 스포츠용품 판매에 관련된 전자우편의 경우, '종류'가 '스포츠' 이고, 용품을 판매하는 '광고의 성질'을 가지는 '범주' 라고 판단할 수 있다. 본 논문에서는 전자우편의 범주를 결정하기 위하여 기본적으로 '광고, 뉴스, 사업, 과학' 등의 몇 가지의 범주를 미리 정의하고 이를 기반으로 스팸메일 필터링을 수행한다. 본 논문에서는 광고의 범주 안에 드는 전자우편을 스팸메일이라고 간주한다. 또한 전자우편의 종류는 '약, 주식, 스포츠, 성, 컴퓨터' 등으로 나눠지며, 같은 종류의 전자우편이라도 서로 다른 범주로 구분 될 수 있다. 예를 들어, 스포츠용품 판매에 관련된 전자우편과 스포츠 뉴스에 관련된 전자우편은 동일한 종류인 '스포츠' 항목에 속하지만, 스포츠용품 판매는 '광고'의 성질을 가지는 범주에 해당되고, 스포츠 뉴스는 '뉴스'의 성질을 가지는 범주에 해당될 것이다.



(그림 3) 텍스트 의미 분석

(그림 3)은 전자우편의 제목, 본문 그리고 첨부된 이미지 안의 텍스트 들을 추출하여 전자우편의 범주와 종류를 결정하기 위한 텍스트 의미분석과정을 나타내고 있다. 일반적으로 텍스트 기반의 스팸메일들은 메일의 제목이나 본문에 광고성 글이 존재하고 있으며, 이미지가 첨부된 스팸메일의 경우, 첨부된 이미지에 광고성 글이 존재한다. 본 연구에서는 전자우편의 제목과 본문의 텍스트들은 파싱하여 수집하고, 첨부된 이미지 안의 텍스트를 FindReader 광학문자판독기[6]를 이용하여 추출한다. 추출된 텍스트들 중 불용어들은 제거한 후, *포터 스테밍* 알고리즘[7]을 통하여 각 단어들의 어근을 추출하는 작업을 수행한다.

시스템은 추출된 모든 텍스트들을 형태가 있는 추상적인 텍스트들인 *추상문자*와 실체가 명확하게 있는 텍스트인 *형태문자*로 분류한다. 추상문자란 동사, 형용사, 부사 즉, '값싼, 비싼, 빠른, 한정, 대출, 판매' 등의 형태가 없는 텍스트들을 뜻하고, 형태문자는 실체가 명확하게 있는 명사로서, '약, 컴퓨터, 제품의 이름'과 같은 단어들을 뜻한다. 시스템은 전자우편 내에 존재하는 추상문자와 형태문자의 빈도수에 따라 해당 전자우편의 종류와 범주를 결정한다. 본 논문에서는 전자우편 안에 존재하는 텍스트를 추상문자와 형태문자로 분류하기 위하여 언어온톨로지인 WordNet[8]을 이용한다. WordNet에 존재하는 텍스트 중 명사는 크게 '사랑, 대출, 추상' 등과 같은 추상적인 명사와 '물건, 물질, 제품' 등과 같은 물질적인 명사로 분류된다. WordNet 에 의해 분류되는 명사 중 추상적인 명사는 추상문자로 결정되고, 물질적인 명사는 형태문자로 분류된다. 또한, WordNet 에 존재하는 텍스트 중 형용사, 동사, 부사 또한 추상문자로 결정한다. 본 논문에서는 이러한 텍스트에 대한 품사를 결정하기 위하여 Java API 를 제공하는 JWordNet[9]을 이용한다. JWordNet 에서 제공되는 WordDataBase 는 각 텍스트들에 대한 품사(명사 = n, 동사 = v, 형용사 = a, 부사 = r)를 정의할 뿐만 아니라, 각 텍스트에 대한 동의어, 반의어와 같은 관련 단어에 대한 검색이 용이하다는 장점이 있다. WordNet 에 의하여 결정된 형태문자와 추상문자는 각각 (그림 2)에서 정의되어 있는 종류와 범주 들 중 하나로 맵핑된다. 본 논문에서는 Jiang 과 Conrath [10]가 제안한 WordNet 상 노드 사이의 거리를 구하기 위한 알고리즘을 이용하여 각 텍스트를 적절한 범주와 종류로 맵핑시킨다. 예를 들어, 범주를 정하기 위한 추상문자 중 '판매, 대출, 값싼, 가격' 등과 같은 텍스트는 '광고'의 범주 안에 속하게 된다. 그리고 종류를 정하기 위한 형태문자 중 '모니터, 프린터, 마우스, 키보드' 등과 같은 텍스트는 '컴퓨터'의 종류에 속하게 된다.

전자우편의 종류를 결정하기 위하여 전자우편에 포함된 형태문자의 빈도수를 계산한다. 식 (1)은 전자우편에 포함된 모든 형태문자의 수(*Nc_terms*)를 구하고, [10]에 의하여 결정된 각 형태문자의 종류컨셉(C)중 그 개수(*Nc*)가 최대인 종류컨셉을 전자우편의 종류(*Text_C*)로 결정한다.

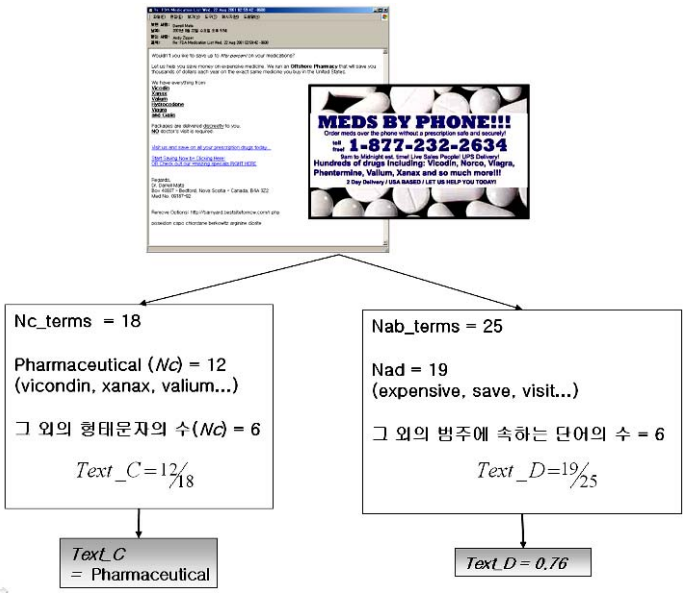
$$Text\_C = \underset{C}{argMax}\left(\frac{Nc}{Nc\_terms}\right) \quad (1)$$

전자우편의 위험도는 해당 전자우편이 광고의 범주에 속할 가능성(스팸 가능성)을 판단하는데 사용된다. 식

(2)는 전자우편에 포함된 모든 추상문자의 수(*Nab_terms*)를 계산하고, 추상문자 중 광고의 범주에 속하는 텍스트들의 개수(*Nad*)를 계산하여 전자우편의 위험도를 분석한다. 시스템은 분석된 값을 전자우편의 텍스트 위험도(*Text_D*)로 결정한다.

$$Text\_D = \frac{Nad}{Nab\_terms} \quad -(2)$$

(그림 4)는 앞서 설명된 식 (1, 2)을 이용하여 전자우편의 종류와 스팸메일의 위험도를 구하는 과정을 보여주고 있다. (그림 4)와 같이, 해당 전자우편의 모든 형태문자 18 개 중 'vicodin, xanax, valium, hydrocodone, viagra, cialis' 등과 같은 'pharmaceutical'에 관련된 형태문자가 12 개로 가장 많이 존재하므로, 'pharmaceutical'이 전자우편의 종류(Text_C)로 결정된다. 위험도 또한 해당 전자우편의 모든 추상문자 25 개 중에서 광고의 범주에 속하는 추상문자 'expensive, save, visit, buy'등 19 개를 통해 텍스트 위험도(Text_D)의 값 0.76 이 결정된다.



(그림 4) 전자우편의 종류와 위험도 분석

3.3. 시각적 의미 분석

시각적 의미 분석 단계에서는 스팸메일에 첨부된 이미지 안의 텍스트와 텍스트의 배경과의 색상의 대조도를 분석하여, 해당 이미지가 가지고 있는 시각적 위험도를 결정한다.

대부분의 스팸메일에 첨부된 이미지 안의 텍스트들은 시각적으로 매우 자극적이고 강렬(빨강, 파랑, 등의 원색)하다. 그리고 첨부된 이미지의 배경은 흰색이나 검정색등의 단순한 원색들로 이루어진 인위적인 배경을 사용함으로써 텍스트의 색상을 더욱 부각시키게 된다. 또한 자연적인 배경 즉, 풍경이나 사물이 있는 배경을 이용한 광고성 이미지들은, 이미지 내에 존재하는 텍스트들의 색상들과 텍스트의 배경색상은 매우 이질적인 경향을 가진다. 즉, 전자우편에 첨부된 이미지내의 텍스트 색상과 텍스트 배경색상의 대조성이 크거나 이질적인 경향을 가지게 되면 해당 전자우편은 스팸메일로 분류될 가능성이 높아지기 때문에, 전자우편에 첨부된 이미지 안의 텍스트 영역 중 텍스트의 색상과 텍스트를 제외한 배경의 색상을 대조해야 할 필요가 있다. 텍스트와 텍스트 배경의 색상을 대조하여 계산된 결과값은 시각적 위험도로 결정된다.

본 논문에서는 Hrishikesh B. Aradhye 의 연구 [4]에서 제안된 알고리즘을 이용하여 이미지 안의 텍스트 범위를 추출한다. 시스템은 이미지 안의 텍스트 범위가 추출되면, 사용된 이미지의 배경이 인위적인지 자연적인지를 파악한다. 보통 자연적인 색상은 32bit 의 색상 즉, 약 42 억 가지의 색상을 표현이 가능하다. 또한 인위적으로 생성된 배경은 색상을 표현하는데 한계가 있어 기본적으로 자연적인 색상과는 많은 차이를 가지게 된다. 시스템은 이미지내의 배경색상의 수에 대한 임계값(T)을 주며, 배경색상이 임계값보다 낮으면 인위적인 배경으로, 높으면 자연적인 배경이라고 판단하게 된다. 본 논문에서는 첨부된 이미지의 텍스트 배경이 자연적인 배경이면 기본적인 시각적 위험도를 부여한다. 반면 첨부된 이미지의 텍스트 배경이 인위적인 배경이라고 판단되면, 시스템은 배경색상과 텍스트 색상에 대한 대조도를 구하여 시각적 위험도를 결정한다. 시스템은 이미지 내의 색상 대조도를 구하기 위하여 RGB 로 표현된 색상을 이미지의 밝기로 표현해야 한다. 본 논문에서는 *국가 텔레비전 시스템 위원회(NTSC)* 표준에 의한 RGB 를 통한 이미지의 밝기를 구하는 공식 식(3)를 이용한다.

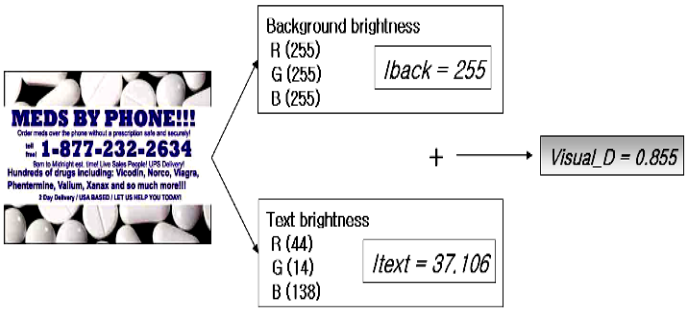
$$brightness = (0.299R) + (0.587G) + (0.114B) \quad -(3)$$

공식 (3)으로부터 얻어진 이미지 내의 텍스트 색상밝기(*Itext*)와 텍스트의 배경 색상밝기(*Iback*)를 이용하여 두 색상간의 대조도를 계산한다. 색상간의 대조도 계산은 수정된 Weber 의 색상 대조 계산공식인 식(4)를 이용한다. 계산된 텍스트 색상과 배경 색상간의 대조도는 시각적 위험도 (*Visual_D*)를 결정하게 된다.

$$Visual\_D = \left| \frac{Itext - Iback}{255} \right| \quad -(4)$$

만약 추출된 텍스트영역의 사용된 배경 색상의 개수가 2 개 이상일 때, 시스템은 배경색상의 사용된 비율을 계산한다. 계산된 배경색상의 비율이 균등하지 않다면 시스템은 사용 비율이 가장 높은 배경색상을 선정하고, 텍스트 색상의 대조도를 계산하여 시각적 위험도를 결

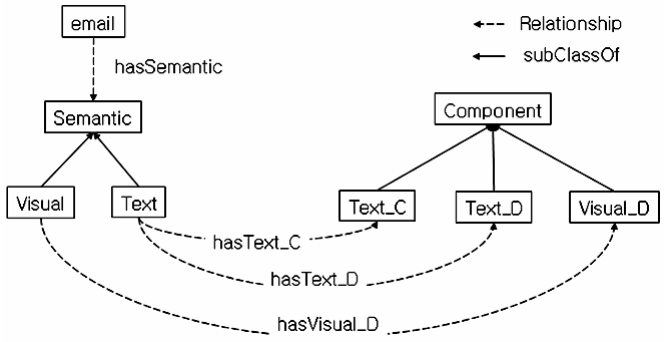
정한다. 그러나 사용된 배경색상의 비율이 균등하면 시스템은 사용된 n 개의 배경색상에 대한 텍스트와의 대조도를 각각 구하고, 대조도가 가장 큰 배경색상을 선정하여 시각적 위험도를 결정한다.



(그림 5) 첨부된 이미지의 색상 대조 값 계산

(그림 5)는 전자우편에 첨부된 이미지를 통해 시각적 위험도를 계산하는 과정을 보여주고 있다. 시스템은 전자우편에 첨부된 이미지로부터, [4]에서 제안된 기법을 이용하여 이미지 안의 텍스트 범위를 추출하고, 해당 범위 내의 텍스트와 배경을 분석한다. (그림 5)에 보여주고 있는 이미지는 텍스트의 색상이 R=44, G=14, B=138 로 이루어져 있고, 텍스트의 배경색상은 R=255, G=255, B=255 로 이루어져 있다. 이를 식(3)에 각각 적용시키면 Itext 의 값은 37.106 을 가지게 되고, Iback 은 255 를 가지게 된다. 시스템은 최종적으로 식(4)를 이용하여 텍스트의 색상의 대조도를 계산하고, 이를 통해 시각적 위험도 (Visual_D)의 값 0.855 가 결정된다.

3.4. 의미추론규칙을 이용한 필터링

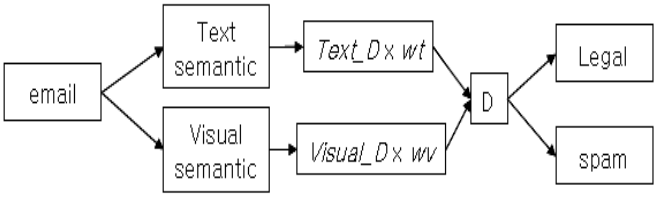


(그림 6) 전자우편 온톨로지

(그림 6)는 시스템에 의해 분석된 전자우편의 정보들을 기술하기 위한 전자우편 온톨로지의 일부분을 나타낸다. 전자우편 온톨로지의 중요 클래스들은 다음과 같다

- **Email 클래스** : 시스템이 스팸메일임을 판별하기 위한 전자우편의 객체
- **Semantic 클래스** : Email 에 존재하는 텍스트, 시각적 의미를 기술
- **Component 클래스** : Semantic 의 분석된 값

예를 들어 email 은 Semantic 과 'hasSemantic' 관계를 가지고 있으며, 또 Semantic 은 각각 Visual 과 Text 를 가진다. Visual 은 'hasVisual_D'의 관계를 가지는 Component 'Visual_D'를 가진다.



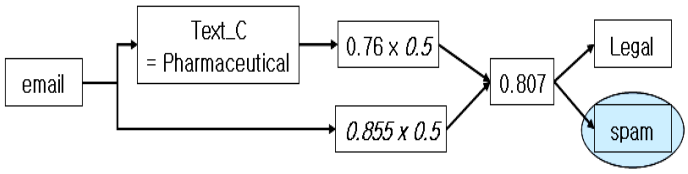
(그림 7) OASIS 의미추론규칙

(그림 7)은 시스템에 의해 생성되는 규칙의 진행과정을 보여주고 있다. 전자우편 온톨로지는 해당 전자우편의 Text_D 를 가지는 Text Semantic 과 Visual_D 를 가지는 Visual Semantic 을 가지고 있다. 각각 생성된 위험도는 전자 우편의 종합 위험도(*D*)를 결정하게 되며 시스템은 전자우편의 종합 위험도를 결정하기 위하여 텍스트 기반의 가중치(*Wt*), 시각적 기반의 가중치(*Wv*)를 부여 한다. 여기서 부여된 가중치는 'Wt + Wv = 1'의 값을 가지게 되며 전자우편의 종류에 따라 변경이 가능하다. 만약 전자우편에 이미지가 첨부되어 있지 않으면 Wv 값은 0 에 가까운 값을 가지며, 반대로 전자우편에 이미지만 첨부되어 있으면 Wt 값을 0 에 가까운 값으로 변경이 가능하다. 종합 위험도는 [*0, 1*] 사이의 값을 가지며, 시스템은 종합 위험도를 임계값(*D_th*)을 기준으로 전자우편이 광고의 범주에 드는지를 결정하고 전자우편의 종류 (*Text_C*)로 해당 전자우편을 분류하게 된다.

```
Rule D is :
        IF email(?x) and
        hasSemantic(?x, ?y) and hasText_Di(?y, ?z) and
        hasVisual(?x, ?a) and hasVisual_Di(?a, ?b) and
        ([?x * Wi] + [?b * Wt]) > D_th
Then emailCategory : advertisement (?x)
```

(그림 8) 부침개로 표현된 의미추론규칙

(그림 8)은 (그림 7)에서 설명된 규칙을 '보쌈'추론엔진[11]에서 제공하는 규칙언어인 '*부침개*'를 이용하여 표현된 추론 규칙을 보여주고 있다. 시스템은 전자우편 온톨로지에 기술된 Component 클래스의 Text_C, Text_D, Visual_D의 값들을 읽어 들여 '보쌈'추론엔진에 적용시킨다. Text_D는 부침개로 표현된 '?z' 값에 삽입되며, Visual_D는 '?b'에 삽입된다. Text_C의 값은 최종적으로 해당 전자우편의 종류를 표현할 때 사용된다.



(그림 9) OASIS 의미추론규칙을 통한 스팸메일 필터링

(그림 9)는 (그림 4,5)에서 설명한 예제를 (그림 7)에 적용하여 실제 필터링이 이루어지는 과정을 보여주고 있다. 해당 전자우편은 텍스트 의미분석에 의해 'Pharmaceutical'라는 Text_C의 값과, 텍스트 위험도(Text_C) 0.76을 가지고, 시각적 의미 분석에 의해 시각적 위험도(Visual_D) 0.855를 가진다. 시스템은 종합 위험도 계산을 위한 Wt, Wv의 가중치를 각각 0.5로 부여를 하였으며, 의미추론규칙에 의해 계산된 종합 위험도는 0.807을 가진다. 시스템에서 광고의 범주에 속할 가능성을 판단하는 임계값 (D_th)을 0.65 이라고 주었다면, 해당 전자우편은 광고의 범주(0.807 > 0.65)에 속하며, Pharmaceutical의 내용을 가지는 전자우편이다. 본 논문에서는 광고의 범주 안에 속한 전자우편을 스팸메일로 판단하므로, 해당 전자우편은 스팸메일로 판단하게 된다.

4. 결론 및 향후 연구

본 논문은 전자우편의 시각적, 텍스트의 의미 정보를 추출하여 온톨로지에 기술하고, 의미추론규칙을 적용함으로써 새로운 기법이 적용된 스팸메일을 효과적으로 필터링 기법을 제안하였다.

제안된 시스템의 장점은 다음과 같다. 첫째, 현존하는 스팸메일의 기법들에 상관없이 전자우편의 시각적, 텍스트의 의미 분석을 하여 스팸메일을 필터링을 할 수 있다. 둘째, 기존의 기계학습을 이용한 필터링기법은 필터링을 위하여 기계학습을 시키는 비용이 큰 것에 비해, 제안된 시스템은 트레이닝을 위한 비용이 적다. 셋째 기계학습이 아닌 의미추론규칙을 사용함으로써, 생성된 필터링 모델에 대한 분석, 관리가 가능하다.

향후 본 논문에서 제안된 시스템의 객관적인 성능 평가를 위하여 약 100 만개의 스팸메일을 이용한 다양한 환경에서의 실험을 수행할 예정이다. 또한 같은 전자우편이라도 사용자 취향에 맞게 분류할 수 있는, 사용자 위주의 환경 적응형 스팸메일 필터링 시스템에 관한 연구를 수행하고자 한다. 예를 들어 '컴퓨터 정보' 에 관련된 전자우편은 컴퓨터 공학자에게는 필요한 전자우편이 되겠지만, 금융업자에게는 스팸메일로 분류가 될 수 있으므로 스팸메일 필터링 시스템은 사용자의 환경에 맞는 분류를 해야 할 필요가 있다. 추가적으로 텍스트, 이미지가 첨부된 스팸메일 뿐만 아니라 동영상과 같은 다양한 미디어를 동원한 새로운 기법의 스팸메일을 대비하여 동영상이나 플래시 등의 시각적 분석을 고려한 연구를 수행할 예정이다.

5. Acknowledgement

이 논문 또는 저서는 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임"(KRF-2006-521-D00457).

참고문헌

[1] Fumera, G., Pillai, I., and Roli, F., Spam Filtering Based On The Analysis Of Text Information Embedded Into Images, *The Journal of Machine Learning Research, 7*, 2699 (2006).

[2] Kim, H. J., Kim, H. N., Jung, J. J., and Jo, G. S., Spam Mail Filtering System Using Semantic Enrichment, *Proc. of the 5 thInternational Conference on Web Information Systems Engineering* (2004).

[3] Kim, J., Dou, D., Liu, H., and Kwak, D., Constructing a User Preference Ontology for Anti-spam Mail Systems, *LECTURE NOTES IN COMPUTER SCIENCE, 4509*, 272 (2007).

[4] Aradhye, H. B., Myers, G. K., and Herson, J. A., Image analysis for efficient categorization of image-based spam e-mail, *Document Analysis and Recognition, Proceedings. Eighth International Conference on*, 914, (2005)

[5] Nhung, N. P., and Phuong, T. M., An Efficient Method for Filtering Image-Based Spam E-mail, *LECTURE NOTES IN COMPUTER SCIENCE, 4673*, 945 (2007).

[6] http://abbyy.com

[7] http://tartarus.org/~martin/PorterStemmer/

[8] http://wordnet.princeton.edu

[9] http://jwn.sourceforge.net

[10] Jiang, J. J., and Conrath, D. W., Semantic similarity based on corpus statistics and lexical taxonomy, *Proceedings of International Conference on Research in Computational Linguistics*, 19 (1997).

[11] Jang, M., and Sohn, J. C., Bossam: An Extended Rule Engine for OWL Inferencing, *Rules And Rule Markup Languages For The Semantic Web: Third International Workshop, RuleML 2004, Hiroshima, Japan, November 8, 2004: Proceedings* (2004).