

정보기술용어의 표준화를 위한 연구

김재호, 류범모, 신호식, 이상석, 최기선

KAIST 전산학과 / 전문용어언어공학센터 / 언어자원은행

{jjaeh, pmryu, gerling, sslee, kschoi}@world.kaist.ac.kr

A Study on the Standardization of IT Terms

Jae-Ho Kim, Pum-Mo Ryu, Hyo-Shik Shin, Sang-Seok Lee, Key-Sun Choi

KAIST CS Dept. / KORTERM / BOLA

요약

지식의 확산과 대중화로 인하여 동일한 개념에 대한 검증되지 않은 용어들이 난립하고 있다. 종전에는 해당 학회의용어심의회위원회, 혹은 전문가 그룹의 합의를 통해서 출판된 전문용어집이 권위를 가지고 전문용어의 사용을 통제하였었다. 그러나 이제는 기존 전문용어집에 수록되지 않은 새로운 개념을 위한 신조어의 생성에 있어서 사용자의 자의적인 해석이 빈번해져서 사회적 합의에 도달하지 못한 새로운 전문용어들이 경쟁적으로 사용되는 추세이다. 이러한 현실에 직면하여 국내외적으로 전문용어의 정비 및 표준화 사업은 언어 및 학문 분야의 경계를 넘어 총체적으로 전략적 접근을 시도하고 있다. 이에 본 연구에서는 7가지의 전문용어 표준화 지침을 수립하고, 그에 따라 ISO 2382의 4030개 정보기술용어를 표준화하고자 한다.

1. 서론

지식의 확산과 대중화로 인하여 동일한 개념에 대한 검증되지 않은 용어들이 난립하고 있다. 종전에는 해당 학회의용어심의회위원회, 혹은 전문가 그룹의 합의를 통해서 출판된 전문용어집이 권위를 가지고 전문용어의 사용을 통제하였었다. 그러나 이제는 기존 전문용어집에 수록되지 않은 새로운 개념을 위한 신조어의 생성에 있어서 사용자의 자의적인 해석이 빈번해지면서 사회적 합의에 도달하지 못한 새로운 전문용어들이 경쟁적으로 사용되는 추세이고, 이러한 용어들이 KS화로 표준화되기도 하고 있다.

ISO 2382군 ISO 2382는 정보기술용어에 대한 국제규격으로 영어와 불어로 되어 있으며, 34개 하위부류에 대해 총 4,030개의 용어를 포함하고 있다. 의 약 4천여 개 용어 중 2,788개만이 KS 용어로 규격화되었다. 아직 규격화되지 않은 용어들에 대해서 KS화가 시급하며, 현재의 KS 우리말 용어에 있어서도 크게 두 가지의 문제가 있다.

첫째, 우리말화를 위한 노력 없이 "다이내미사이저 (serializer)", "세션 (session)", "슬라이싱 (slicing)", "스테티사이저 (staticizer)" 와 같이 외국어를 음차표기한경우가 너무 많다. 음차표기 방식은 용어의 개념이 투명하게 드러나지 않으며, 용어의 난해성을 가중시켜 학습의 비효율성을 가져온다. 둘째, 용어의 일관성에 문제가 있다. "directory"는 독자적인 용어로서 혹은 용어의 성분어휘로서 사용되는데, "디렉토리 시스템 대리인 (directory system agent)", "등록부 사용자 대리인 (directory user agent)" 과 같이 일관성 없이 우리말로 대역되고 있다. 이외에도 KS의 우리말 용어는 일본식 한자어 등을 통해 어려운 용어가 많다는 점에서도 용어의 사용을 어렵게 하는 문제가 있다.

용어 표준화를 위한 연구로 김진용 [1]의 연구가 있었으나, 표준화 방법 수립을 위한 자료 분석에 머무는 수준이었다. 본 연구에서는 ISO 704 [2] 및 그에 근거한리수락 [3]의 "용어 표준화 원칙"을 참고하여

전문용어의 표준화 지침을 수립하고, 그에 따라 ISO 2382의 4,080개 정보기술용어를 표준화하고자 한다.

2. 용어 표준화 원칙

본 연구에서 수립한 7가지의 표준화 지침은 다음과 같다.

[제 1조: 원어와 대역어의 수준 일치화] 용어는 되도록 쉬운 어휘를 사용하여 구성한다.

ISO 영어 용어가 쉬운 용어이던 대역어로서 한국어 용어도 쉽게 하여야 한다. 새로운 전문용어를 만드는 방법은 이미 쓰고 있는 전문용어들을 합성하기도하고 일상 생활 용어나 다른 분야의 전문용어를 빌려 쓰기도 한다. 예를 들어, 합성 전문용어로는 "information retrieval" ("정보검색")을 들 수 있고, 일상 생활 용어에서 차용한 것으로는 "window" ("창")를 들 수 있다.

용어를 형성하는데 사용된 성분 어휘들에 대해서 가능한 일상적인 어휘 표현인 기초어휘를 선택함으로써 궁극적으로 쉬운 용어를 표준안으로 택한다. 이러한 배경에는 용어를 학습하는 사람들의 단계별 수준에 용어의 난이도가 부응해야 한다는 인식이 깔려 있다. 예를 들어, 초등학교 교재에 등장하는 용어는 초등학교 어휘력 수준에서 개념적으로 이해될 수 있는 것이어야 한다. 단, 영어 용어의 어휘 자체가 매우 어려운 수준일 경우에는 무리하게 쉬운 용어로 바꾸지 않아도 된다.

예) "parameter"의 우리말 대역어는 "매개변수", "가인수" 모두 가능하지만 "매개변수"의 등급(4등급 본 연구에서는 어휘의 난이도 결정하기 위하여 KORTERM의 기본어휘[5] 등급표를 이용한다. 1~5등급까지 있으며, 1등급이 가장 기본이 되는 어휘이다. 등급 외 어휘를 본 연구에서는 6등급으로 취급한다.)이 "가인수"의 등급(6등급)보다 낮은 쉬운 어휘이기 때문에, "parameter"는 "매개변수"로 대역된다.

[제 2조: 순우리말화] 외국어 음차표기 보다는 순우리말 용어를 받아들인다.

"외국어 단어를 모국어로 옮길 때, 발음을 빌어 모국어 글자로 적는 음차표기는 가장 쉽지만 가장 게으른 방법이다. 이런 방법을 이용할 경우 모국어의 자연스러운 언어환경과 단절되어 발전성이 없다" [4]. 단순히 음차표기 방식을 빌린 외국어의 한글 표기를 우리 용어로 받아들이는 경향이 점점 높아지고 있는 실정이다. 다른 용어 성분으로 포함되어 많이 사용되는 기본 용어까지도 외국어 음차 표기를 하면서 외국어 음차표기가 범람하게 되고, 용어에 대한 개념적 이해가 어렵게 되었다. 이러한 문제를 해결하기 위해서 가급적이면 이미 고착화된 외래어 용어들에 대해서도 순우리말화 노력을 기울여야 한다.

이미 보편적으로 많이 쓰여 굳어진 외래어 용어에 대해서도 고유어 혹은 쉬운 한자 용어를 만들어 복수 표준안을 제시함으로써 경쟁을 시키도록 유도한다. 고유어를 우선적으

로 받아 들이되, 표현에 있어서 파생력에 문제가 있거나 다른 어휘와 결합하여 새로운 용어를 만드는데 어색함이 있을 경우에는 한자어를 사용한다 한자어를 사용할 경우에는 가급적이면 쉬운 한자어를 사용하도록 한다.

예) "checkpoint"는 "체크포인트", "검사점" 모두 가능하지만, 고유어 또는 쉬운 한자어 우선 원칙에 따라 "검사점"으로 한다.

예) "icon"은 "아이콘"으로 굳어져 버렸기 때문에, 고유어/쉬운 한자어인 "그림기호"를 "아이콘(=그림기호)"에 복수 표준안으로 추가하여 경쟁하도록 한다.

[제 3조: 일관성] 용어 및 용어 내 성분 어휘는 일관성을 갖도록 한다.

개념상 동일한 원어 성분 어휘에 대해서 일관되게 우리말 대역어를 선택하도록 한다. 이들 기본 어휘들은 반복적으로 많은 용어에 성분 어휘로 사용되기 때문에 용어의 일관성을 획득하는데 파급효과가 크다. 기본 어휘들의 단일화를 통해서 영어 원어에 대한 우리말 용어의 일관성을 유지하도록 한다. 반대로 우리말 용어에 대한 영어 원어도 일관성을 유지하도록 한다.

예) "assignment"는 "대입", "할당", "대입", "지정" 등이 가능하지만, "배정"으로 선택하였다고 하자. 그렇다면 의미상 문제가 없을 경우에만하여 "assignment"가 들어간 모든 용어에서 "assignment"는 "배정"으로 번역되도록 한다. 즉, "assignment by name"은 "이름방식 배정", "assignment statement"는 "배정문", "credit/blame assignment"는 "신용/책임 배정"으로 한다.

[제 4조: 정확한 개념표현] 용어의 개념을 가장 잘 표현하는 용어를 선택하도록 한다.

용어의 정의문 정보를 참조하여 용어의 개념을 잘 드러낼 수 있는 표현을 선택하도록 한다. 다만, 용어의 개념적 정의를 그대로 번역하여 우리말 대역어를 제시하는 방법을 택하기 보다는 용어를 확장적으로 해석함으로써 해당 개념을 유추할 수 있는지를 주요한 기준으로 삼는다.

(예) "gray scale"의 정의는 "흰색과 검은색 사이의 농도의 범위"이다. "농도 계조", "명암 단계", "그레이 스케일" 등으로 번역이 가능하며, "scale"을 "등급"으로 하는 것이 "농도의 범위"라는 개념을 가장 잘 나타낸다. 그래서 "gray scale"의 대역어로 "회색 등급"을 선택한다.

[제 5조: 고빈도 선호] 말뭉치상의 빈도수가 높은 용어를 우선적으로 선택한다.

이 기준은 위 상위 4개([제 1조], [제 2조], [제 3조], [제 4조])의 원칙에 의해 생성된 용어 후보들이 경쟁적일 때 사용하는 보조적인 원칙이다.

이 기준은 용어의 생명력과 관련이 있다. 현재 널리 사용되고 있는 용어를 표준으로 선정하려는 데 목적이 있다. 여

러 가지 후보 대역어 중 말뭉치에서 가장 많이 나타나는 용어를 선정한다. 말뭉치의 크기가 작아서 통계적으로 의미 있는 빈도수를 구하기 어려운 경우에는 웹 검색 엔진에서의 빈도수를 이용할 수 있다.

예) "associative learning"의 우리말 대역어로 "상관학습", "연관학습" 등이 가능하지만 웹에서 가장 많이 검색된 "연관학습"을 대역어로 선정한다.

[제 6조: 사전 등록 용어 선호] 기존 용어 사전에 등재된 용어를 우선적으로 선택한다.

기존의 정보통신 분야 사전에 등재여부를 이용하여 표준 용어를 선정한다. 기존의 사전에 등재된 경우는 그 용어가 널리 통용되고 있다는 사실을 의미하기 때문에 기존 사전에 등재 여부를 표준 용어의 기준으로 삼는다. [제 5조: 고빈도 선호]의 원칙과 마찬가지로 이 기준도 역시 후보들이 경쟁적일 때에 적용하는 보조적 원칙이다.

예) "interoperability"는 "상호운용성", "호환성", "상호연동성"으로 번역될 수 있지만 "상호운용성"이 다수의 사전에 등록되어 있기 때문에 "상호운용성"을 대역어로 선정한다.

[제 7조: 기능어 삭제] 기능어는 되도록 삭제한다.

용어의 언어적 경제성의 원칙에 따르면 용어는 가능하면 간략하여야 한다. 따라서용어가 내포하는 개념을 충분히 전달할 수 있는 경우에는 용어에 사용되는 기능어를 삭제한다. 용어에서 많이 사용되는 기능어는 관형격 조사 "-적", "-의" 등이 있다.

예) "logical ring"은 "논리적 고리", "논리 고리"로 번역될 수 있지만 의미 전달에 애매성이 없기 때문에 용어의 경제성 원칙에 따라 "논리 고리"를 대역어로 선정한다.

3. 용어표준화 작업 및 결과

본 연구에서는 2장에서 세운 표준화 심의 지침을 ISO 2382의 4,030개 2001년 남북합의안 민간 학술단체 차원에서 수행된 것으로 당시 남한 내부의 표준안 작업이 미흡하여, 대부분 북한안이 남북 합의안으로 채택되었다. 그 결과 기존의 남북 합의안은 실제 사용에 있어서 파급효과가 적었다. 용어에 적용하여 새로운 표준용어를 제안하였다. 표준용어는 다음과 같은 4단계의 절차를 거쳐 만들어진다.

1단계: 용어의 조어분석

4,030개의 모든 한국어 용어를 조어 분석하여 영어-한국어 성분어휘 쌍을 구축한다. 각 영어 성분어휘 별로 한국어 대역어를 표준화하여, 개념의 차이가 없는 한 각 성분어휘들이 일관성 있게 원어에 대해서 우리말로 대역하도록 하기 위해서이다. 총 4,030건의 용어에 대해서 7,587개의 성분어휘 쌍이 구축되었다.

2단계: 우선 심의 성분어휘 선정

심의작업의 효율성을 위해서 중요하다고 생각되는 어휘를 우선적으로 선정한다. 앞서서구축한 성분어휘 쌍 중 여러 용어에 성분어휘로 쓰이는 고빈도 영어 어휘 (102건)와 한국어 대역어가 영어의 음차표기 그대로 사용되는 기초적인 외래어 (304건)는 정비의 파급효과가 크다는 점에서 우선적으로 심의하기 위해서 별도로 분류하였다.

3단계: 우리말 용어의 성분어휘에 대한 대역어 심의

2단계에서 분류된 성분어휘 목록을 토대로 우리말 대역어를 심의하였다. 이 과정에서 용어 표준화 심의 지침이 적용되었다.

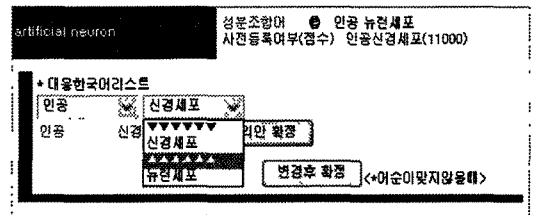
4단계: 용어 차원에서 우리말 용어의 심의

4,030개의 실제 용어를 심의한다. 용어에 3단계에서 마련된 영어성분어휘가 포함된 경우, 그 대역어로 대체하여 한국어 용어를 생성하고 심의한다.

성분어휘들의 합성을 통해서 용어를 생성하면 부자연스런 결합이 될 수도 있다. 따라서, 성분 어휘들 차원에서 단일성을 기본적으로 추구하지만 용어 차원에서는 어느 정도 단일성의 기준을 완화하거나 성분어휘의 대역어 목록을 수정한다. 특히 우리말고유어의 경우에는 성분어휘 차원에서는 좋은 대역어가 될 수 있었지만, 다른 성분어휘들과 결합하는 용어 차원에서는 자연스러움 측면에서 문제가 있는 경우가 있었다. 따라서, 이런 경우에는 차선책으로 다른 성분어휘들과 결합이 자연스럽고 파생력이 높은 쉬운 한자어로 대체하였다.

본 용어의 심의에 있어서는 웹에서 작업할 수 있는 도구를 개발하여 이용하였다. 이를 통해서 용어 심의의 일관성을 유지하고, 여러 사람들의 동시적인 작업 결과가 수시로 갱신되어 중복되지 않고 상호 참조하면서 작업할 수 있도록 편의를 도모하였다.

이 작업 도구는 <그림 1>과 같이 각 영어 성분어휘별로 한국어 대역어 후보를 자동으로 제시해 준다.



<그림 1> 'artificial neuron' 의 성분어휘 별 대역어가 "인공"과 "신경세포"로 자동으로 제시되는 화면

기존 용어의 대역어와 이미 심의를 본 성분어휘의 대역어를 참고하여 제시한다. 제시한 후보 외의 다른 대역어를 넣고 싶거나 어휘의 순서가 바뀌는 경우를 고려하여 선택이 아닌 직접 입력도 가능하도록 되어 있다.

그리고 이 작업 도구는 다른 여러 사전에서의 용어 대역어 목록과 정의문 등의 부가 정보도 모두 볼 수 있게 제공해 준다. 또한, <그림 2>와 같이 용어의 구성어휘성분이 다른 용어에서 어떻게 쓰이고 있는지를 보여 주어 대역어가 올바르게 붙어 사용되고 있는지 알 수 있게 한다.

artificial 인공	artificial neural network ANN 인공 신경 망	인공 신경 망
artificial 인공	artificial intelligence AI 인공 지능	인공 지능
artificial 인공	artificial intelligence(2) AI 인공 지능	인공 지능
artificial 인공	artificial intelligence(1) AI 인공 지능	인공 지능
artificial 인공	hierarchical artificial neural 계층화 인공 신경 망	계층화 인공 신경 망
neuron 신경세포	output neuron 출력 신경세포	출력 신경세포
neuron 신경세포	input neuron 입력 신경세포	입력 신경세포
neuron 신경세포	artificial neuron 인공 신경세포	인공 신경세포
neuron 신경세포	visible neuron 보이는 신경세포	보이는 신경세포
neuron 신경세포	hidden neuron 숨은 신경세포	숨은 신경세포

<그림 2> 용어별 영어성분 목록 전체 보기

4. 용어 표준화의 평가

표준화된 용어는 표준화 원칙에 충실한 정도를 계산하여 정당성을 평가한다. 평가방법에서 2장에서 정리한 용어 표준화 원칙을 충실하게 반영하는 용어에 높은 점수를 부여하고, 반대의 경우에는 낮은 점수를 부여한다. 용어의 점수는 각각의 기준으로 독립적으로 평가한 후, 전체 점수로 합한다. ISO2382 용어의 기존 한국어 대역어와 새로 구축한 표준안 대역어 목록을 비교 평가한다.

[제 1조: 용어의 수준 일치화]에 대한 평가는 두 가지 방향에서 생각할 수 있다. 첫째, 영어 용어의 구성 어휘와 우리말 용어의 구성 어휘가 비슷한 수준에서 결정되어야 한다. 예를 들어 "personal computer"의 우리말 대역어가 "개인용 컴퓨터"인 경우에 영어에서 "personal"와 "computer"의 등급이 각각 우리말에서 "개인용"과 "컴퓨터"의 등급과 비슷하여야 한다. 이를 평가하기 위하여 먼저 영어와 우리말 어휘들을 각각 난이도에 따라서 6등급으로 분류하고 그 차의 평균을 이용하여 계산한다. 둘째, 우리말 용어는 가능하면 쉬운 어휘로 구성되어서 일반인들도 쉽게 용어의 개념을 이해할 수 있어야 한다. 각 어휘의 등급의 평균을 이용하여 계산한다.

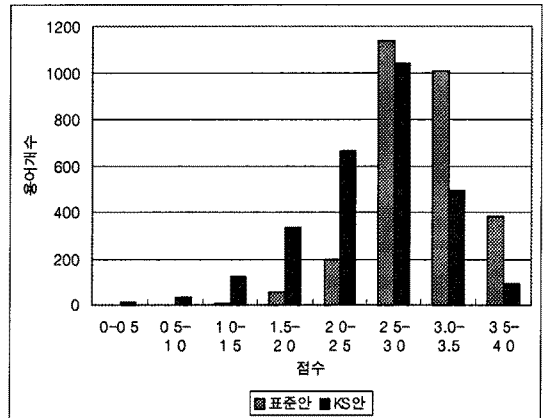
[제 2조: 순우리말화]의 기준에서는 외래어나 한자어보다 우리말 어휘로 표현한 경우에 높은 점수를 준다.

[제 3조: 일관성]에 의하면 동일한 영어 어휘에 대해서 일관되게 단일한 우리말 대역어를 선택하도록 한다. 특히 기본 어휘들은 반복적으로 많은 용어에 성분 어휘로 사용되기 때문에 용어의 일관성을 획득하는데 파급효과가 크다. 기본 어휘들에 대한 우리말 대역어를 단일화하면 영어 용어에 대한 우리말 용어의 일관성을 유지할 수 있다. 먼저 우리말 용어의 성분어휘 단위로 일관성 점수를 계산한 뒤 용어의 일관성 점수를 계산한다. 일관성 점수는 우리말 대역어의 구성 어휘가 영어 용어의 구성 어휘에 대한 일관성 대역어로 구성될수록 1에 가까운 점수를 가지고, 그렇지 않은

경우 0에 가까운 점수를 가진다.

[제 4조: 정확한 개념표현]의 기준을 평가하기 위하여 전문가의 판단에 따른다. 다수 명의 전문가에게 용어의 정의문과 용어를 제시하고, 정의문의 의미의 정확한 표현 여부를 판단하도록 한다. 실험에서 100개의 ISO 용어를 임의로 선택한 후, 표준안과 KS안 용어의 정의문을 각각 2명의 정보통신분야 전문가에게 제시한다. 평가에 참가한 전문가들은 제시된 용어 쌍(표준안-KS안) 중에서 주어진 정의문의 의미를 잘 전달하는 용어를 선택한다. 전문가들은 주어진 용어들에 대한 표준안 여부를 알지 못한 상태에서 정의문의 개념을 명확히 전달하는 용어를 선택한다. 개념을 잘 전달한다고 선택된 용어 중에서 표준안 용어가 많이 포함된 경우는 표준안이 더 우수하다고 할 수 있고, 그렇지 않은 경우는 다른 표준안이 우수하다고 할 수 있다.

[제 1조]의 두 항목, [제 2조], [제 3조], [제 4조]의 점수를 각각 0~1 사이의 값을 가지도록 정규화하였으며, 이를 합산하여 모든 용어의 표준안과 KS안에 대한 점수를 계산하였다. 따라서 모든 용어는 0점에서 5점 사이의 점수를 가진다. <그림 3>은 표준안과 KS안 용어들의 점수대별 분포를 보여준다. 표준안에서 2.5점 이상의 용어가 90.7% (2,526개)이고, KS안에서 2.5점 이상의 용어가 58.2% (1,620개)였다.



<그림 3> 표준안과 KS안 용어들의 점수대별 분포

다음 <표 1>과 <표 2>는 본 연구에서 제안한 표준화지침의 의거하여 표준화한 용어의 예들이다.

<표 1> 순우리말화한 용어 목록

ISO 용어	표준안		기존 KS 용어
	제1표준안	동의어	
alphabet	알파벳	자모	알파벳
backup	백업	여벌	백업

cartridge	카트리지	교체통	카트리지
debugger	디버거	벌레잡개	
alias	별칭	에일리어스	
ring	고리	링	
catalog	목록화	카탈로그	카탈로그, 목록

분에 대해 계속적으로 표준화를 확대 적용해야 하며, 다른 한편으로는 마구 쏟아지는 신조용어들에 대해서 현 표준화 방향에서 일관성 있게 적용해야 하는 양방향의 과제가 남아 있다.

감사의 글

본 연구는 "남북 정보산업용어 표준(안) 도출을 위한 기초 연구"의 일환으로 한국산업자원부/표준협회의 지원을 받았습니다. 특히, 용어 표준화 지침을 마련하고 용어를 심의하는데 정보 기술 관련 각 분야 전문가들로 구성된 심의위원회의 활동이 컸습니다.

참고 문헌

- [1] 김진용, 이병희 (2003), "동일 어종 내의 용어 조화 세부 원칙에 대하여", 제6회 동아시아전문용어포럼 (EAFTerm) 발표 논문집.
- [2] ISO 704: 2000 (E) Terminology work: Principles and methods
- [3] 리수락 (2001), "정보기술 용어의 우리말 다듬기와 표준화에 관한 연구", 초록

<표 2> 일관성을 고려하여 표준화한 용어 목록

단일어	ISO 용어	표준안
delivery (배달)	delivery notification service	배달 통지 서비스
transmission (전송)	text transmission	본문 전송
transfer (이송)	message transfer system	메시지 (=통신문) 이송 체계

5. 결론 및 기대효과

본 연구에서는 한편으로는 전문용어의 표준화 지침을 마련하고, 다른 한편으로는 구체적으로 ISO 2382 정보기술분야 용어를 중심으로 표준안을 마련하고자 하였다.

기존 우리말 용어들의 분석을 통해서 표준화 지침을 마련하였지만 모든 용어에 대해서 이러한 지침을 충분히 적용했다고는 볼 수 없다. 굳어진 외래어 등이 이에 해당하는데 이는 표준안을 도출하기 위해서는 "단칼에 전부를 해결하기"보다는 점진적인 추진전략이 필요함을 의미한다. 사실 현재의 표준화 지침을 통한 용어의 표준안은 현 상황에서 가능한 최대한의 작업결과 일뿐, 향후에는 현재 미흡했던 부

- [4] Wright, Sue-Ellen & Gerhard B'udin (1997), Handbook of Terminology Management, Vol. 1. J. Amsterdam/Philadelphia: Benjamins.

- [5] 배희숙, 이주호, 시정곤, 최기선, 2003, "자연언어처리용 전자사전을 위한 한국어 기본어휘 선정", 언어와 정보, 한국언어정보학회, p41-54.

분에 대해 계속적으로 표준화를 확대 적용해야 하며, 다른 한편으로는 마구 쏟아지는 신조용어들에 대해서 현 표준화 방향에서 일관성 있게 적용해야 하는 양방향의 과제가 남아 있다.