

NCS 능력단위 요소와 기존 교육과정 간 갭 분석을 위한 평가모델

Evaluation Model for Gab Analysis Between NCS Competence Unit Element and Traditional Curriculum

김 대 경¹ · 김 창 복^{2*}

¹에듀엔컴(주) NCS 공학연구소

²가천대학교 에너지 IT학과

Dae-kyung Kim¹, Chang-Bok Kim^{2*}

¹Department of Engineering Research, Education and Company Co., Ltd, Seoul 137-837, Korea

²Department of Energy IT, Gachon University, Gyeonggi-do 461-701, Korea

[요 약]

국가 직무능력 표준 (NCS; national competency standards)은 직무를 수행하기 위해 요구되는 능력에 대한 체계화 및 표준화이다. NCS는 특정 직무능력인 능력단위 요소로 구체화하고 표준화하여 학습모듈을 개발한다. 기존 교육과정은 NCS 능력단위 요소를 교육 훈련에 활용하기 위해서 갭 분석 (gab analysis)이 필수적이다. 기존에 갭 분석은 전문가가 주관적으로 평가하였다. 전문가에 의한 갭 분석은 심리적 요소에 의해 주관적 결정, 정확성 결여, 시간 및 공간적 비효율성 문제가 제기되었다. 본 논문은 주관적 평가의 문제 해결을 위해 자동화 평가모델을 제시하였다. 본 논문은 기존 교육과정과 능력단위 요소 간 갭 분석을 위해, 색인어 추출, 단어빈도수-역 빈도수 기반 특징 값 추출, 코사인 유사도 알고리즘을 이용하였다. 또한, 기존 교육과정과 NCS 능력단위요소 사이 유사도 매핑 테이블을 제시하였다. 본 논문의 평가모델은 구조적 특징이나 속도 면에서 개선된 알고리즘을 통해 보완해야 한다.

[Abstract]

The national competency standards (NCS) is a systematize and standardize for skills required to perform their job. The NCS has developed a learning module with materialization and standardize by competence unit element, which is the unit of specific job competency. The existing curriculum is material to gab analysis for use in education training with competence unit element. The existing gab analysis has evaluated subjectively by experts. The gab analysis by experts bring up a subject subjective decision, accuracy lack, temporal and spatial inefficiency by psychological factor. This paper is proposed automated evaluation model for problem resolve of subjective evaluation. This paper use index term extraction, term frequency-inverse document frequency for feature value extraction, cosine similarity algorithm for gab analysis between existing curriculum and competence unit element. This paper was presented similarity mapping table between existing curriculum and competence unit element. The evaluation model in this paper should be complemented by an improved algorithm from the structural characteristics and speed.

Key words : National competency standards, Gap analysis, Cosine similarity, Term frequency, Inverse document frequency.

<http://dx.doi.org/10.12673/jant.2015.19.4.338>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 15 July 2015; Revised 24 July 2015

Accepted (Publication) 10 August 2015 (30 August 2015)

*Corresponding Author ; Chang-Bok Kim

Tel : +82-10-8908-3946

E-mail : cbkim@gachon.ac.kr

1. 서론

정부는 오래전부터 직업 교육, 훈련, 자격 제도를 산업현장 중심으로 혁신하고, 기업의 능력 중심 인사관리를 유도하기 위해, 국가 직무능력 표준(NCS; national competency standards)을 개발해 왔다. 그리고 최근에 현 정부는 NCS를 활용하기 위한 정책을 활발하게 추진하고 있다[1],[2]. NCS는 특정 직무에 필요한 능력을 NCS 분류체계의 하위단위이며 기본 구성요소인 NCS 능력단위 요소로 구체화 하고 표준화하여 NCS 학습모듈을 개발한다. NCS 학습모듈은 NCS 능력단위 요소를 교육훈련에서 학습할 수 있도록 구성된 학습 자료이다[3].

NCS 능력단위 요소는 교육 및 훈련현장에서 활용하기 위해서, 기존 교육과정 자체를 사용하는 방법, 기존 교육과정을 능력단위 요소로 전체 대체하는 방법, 기존 교육과정과 능력단위 요소 사이의 유사도 분석을 통해 부분적으로 기존 교육과정을 능력단위 요소로 대체 하는 방법 등으로 전략을 수립할 수 있다. 특히, 마지막 방법인 기존 교육과정과 능력단위 요소 사이의 유사도 분석은 갭 분석(gap analysis) 형태로 이뤄지는데, 지금까지는 NCS 직무 교육 전문가, 직무 내용 전문가, 직무 분석 전문가가 직접 갭 분석을 해야 했다. 전문가에 의한 갭 분석 방법은 심리적 요소에 의해 주관적 결정, 정확성 결여, 시간적, 공간적 비효율성 문제가 제기되었다. 따라서 기존 교육과정과 능력단위 요소 사이의 갭 분석 자동화가 필요하다.

본 논문에서는 기존 교육과정과 능력단위 요소 간 갭 분석의 자동화를 위한 평가모델을 제시하였다. 평가모델은 기존 교육과정과 능력단위 요소 문서에 대해 형태소 분석과 명사 추출을 통해 색인어를 추출하였으며, 단어빈도수(term frequency) - 역빈도수(inverse document frequency) 알고리즘을 이용하여 각 색인어에 가중치를 부여하였다. 또한, 가중치가 부여된 색인어에 대해서 문서 유사도를 효율적으로 측정할 수 있는 코사인 유사도(cosine similarity) 알고리즘을 이용하여, 능력단위 및 기존 교육과정 간에 유사도를 측정하였으며, 각 유사도에 대해 매핑 테이블을 통해 유사도 평가 결과를 도출하였다. 본 논문은 2장에 관련연구로서 NCS 기반 교육과정 개발 방법과 문서 비교 기법에 대해서 서술하였으며, 3장에서는 갭 분석을 위한 평가모델을 제시하였다. 또한, 4장에서 제시한 기법에 대해 실험결과를 보였으며, 최종적으로 5장에서 결론을 내렸다.

II. 관련연구

2-1 NCS 기반 교육과정 개발 방법

NCS는 산업현장에서 직무수행을 위해 요구되는 지식, 기술, 소양 등을 정부가 산업부문별, 수준별로 체계화 한 것으로, 성공적으로 직무를 수행하기 위해 직무능력을 표준화 한 것이다 [4].

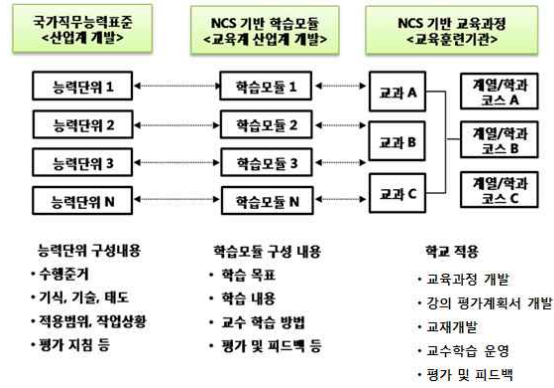


그림 1. NCS 학습모듈
Fig. 1. NCS learning module.

NCS 학습모듈은 능력단위 요소를 교육훈련에서 학습할 수 있도록 구성된 교수-학습 자료이며, 구체적 직무를 학습할 수 있도록 이론 및 실습과 관련된 내용을 상세하게 제시하고 있다. 그림 1에 산업계, 교육계, 교육기관에서의 NCS 학습모듈 개발과 사용 방법에 대해서 나타냈다[5],[6].

NCS 학습모듈은 능력단위 요소 1개당 1개의 학습모듈 개발을 원칙으로 한다. 그러나 필요에 따라 고용단위 및 교과단위를 고려하여 능력단위 몇 개를 묶어서 1개의 학습모듈로 개발할 수 있으며, 능력단위 1개를 여러 개의 학습모듈로 나누어 개발할 수도 있다[6].

NCS 기반 교육과정 개발방법은 교육과정 개편방향 수립, 교육과정 개편, 교육과정 실행방안 수립 등으로 구분된다. NCS 기반 교육과정 개편방향 수립은 외부환경 분석, 내부환경 분석을 거쳐 학과 인력 양성 유형을 설정한다. NCS 기반 교육과정 개편은 직무분석과 NCS 기반 교과목을 도출하여, 교육과정 로드맵, 교과목 프로파일, 강의계획서 등을 작성하게 된다. NCS 기반 교육과정 실행방안 수립은 교수자 역량개발, 교재개발, 교육시설 확보, 산학 협력 현황 등을 실시하게 된다.

여기서 NCS 기반 교과목 도출을 위해서는 현행 교육과정과 유사 능력단위를 선택하여, 다양한 갭 분석과 능력단위 매트릭스(matrix)를 작성하여야 한다. 그림 2에 NCS 학습모듈에 기초한 교육과정 개발 과정에 대해서 나타냈다.

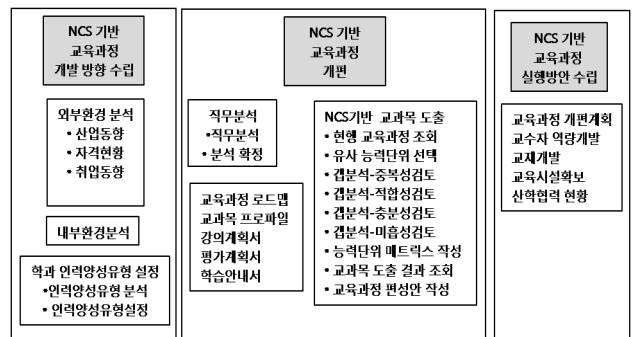


그림 2. NCS 기반 교육과정 개발 과정
Fig. 2. NCS based on curriculum development process.

표 1. 문서비교 기법

Table 1. Document compare technique.

문서비교 기법	방법	특징
어절트리 비교	빈도수 어절 색인을 이용해 색인어 간의 유사도 비교	단어 빈도의 유사도에 의존
패턴 매칭 방식	프로그램 소스를 패턴화시켜 하나의 언어로 매칭하여 탐지	탐지시 조건이 제한적
지문법 이용	문서의 단어들의 유사성, 단어의 빈도수를 비교하여 통계적으로 분석	유사어 사용, 문법을 변경할 경우 탐지 불가. 문서량 증가시 많은 시간 소요.
N-gram 방식	N-gram의 음절을 추출 및 비교	많은 저장 공간 필요. 전혀 다른 문장도 유사한 것으로 판정가능
벡터 공간 모델 방식	문장에서 색인어를 추출하여 공간상의 벡터로 표현해서 유사도를 계산	추출되는 키워드가 정확히 일치해야함
LSA 방식	단어들의 의미 관계를 찾아 유사 여부 판별	정확도가 떨어짐

2-2 문서비교 기법

문서비교 기법은 문서 간 유사도 정도를 측정하여, 문서 표절, 문서 비교, 문서 유사 정도를 나타내는 기법이다. 표 1에 기존 문서 비교 기법에 대해서 나타냈다[7].

한글 문서비교를 위해서는 구문태깅 및 형태소 분석 그리고 명사추출을 통해 색인을 추출하여, 각 색인별 가중치를 계산하는 단계가 필요하다. 색인(index)추출은 단어 색인과 단어 쌍 색인 추출이 있다.

단어 색인의 경우 명사 위주의 단일 단어로 구성되어 적은 양의 데이터에서도 많은 색인을 추출할 수 있다. 그러나 단어 색인은 단어 간 직교성을 지닌 정보로서, 문맥(context) 정보를 포함할 수 없다는 단점이 있다.

단어 쌍 색인은 단어색인 대상으로 동일문장에서 단어를 조합하여 색인어로 추출한다. 단어 쌍 색인은 공기(co-occurrence) 정보를 이용해 문맥 정보를 반영할 수 있다. 공기정보는 두 단어가 동일문서, 문장, 구 등에 같이 발생하는 현상을 말하며, 자주 발생할수록 두 단어가 밀접한 관계를 가지고 있다는 전제에 기반하고 있다[8],[9].

각 단어 및 단어 쌍은 문서에서 사용되는 중요도에 따라, 가중치를 가지며, 문서 비교를 위해서는 이를 수치로 나타내어야 한다. 단어 및 단어 쌍의 가중치를 나타내는 방법으로 단어 빈도수와 역 문서 빈도수 방법이 있다.

단어 빈도수 기법은 문서의 각 단어 및 단어 쌍에 대한 빈도수를 가중치로 이용하는 방법이다. 단어 빈도수 기법은 단어에 대한 빈도수를 그대로 나타내어 사용하게 되면, 연관성 평가에 다소 문제가 발생하기 때문에 다양한 함수를 사용하여 가중치를 보완한다. 표 2에 단어 빈도에 따른 다양한 가중치 공식을 보여주고 있다[10].

표 2. 단어 빈도 가중치

Table 2. Term frequency weight.

함수	가중치 공식
이진함수	$w(k_{ij}) = \begin{cases} 1 & \text{if } f(k_{ij}) > 1 \\ 0 & \text{otherwise} \end{cases}$
단순함수	$w(k_{ij}) = f(k_{ij})$
로그함수	$w(k_{ij}) = 1 + \log f(k_{ij})$
더블로그함수	$w(k_{ij}) = 1 + \log(1 + \log f(k_{ij}))$
루트 함수	$w(k_{ij}) = \sqrt{f(k_{ij})}$
보정함수	$w(k_{ij}) = (1 - \alpha) + \alpha \times \frac{f(k_{ij})}{\max f(k_{ij})}$ 단 $\alpha = 0.5$ 또는 0.6
오가피 함수	$w(k_{ij}) = \frac{f(k_{ij})}{2 + f(k_{ij})}$
더블로그2함수	$w(k_{ij}) = 1 + \log_2(1 + \log_2 f(k_{ij}))$
루트직선함수	$w(k_{ij}) = \frac{f(k_{ij}) + 3}{4}$

$f(k_{ij})$ 는 i 문서에 색인어 j의 빈도수를 나타내며, $w(k_{ij})$ 는 i 문서에 j의 가중치를 나타낸다. 표 2에서 단순함수 $w(k_{ij}) = f(k_{ij})$ 를 사용하면, 기본적인 단어 빈도수이며, 매우 유사한 문서임에도 큰 차이가 발생할 수 있기 때문에, 로그 및 루트 그리고 다양한 변형식을 이용하여 사용한다. 그러나 단어 빈도수로만 가중치를 고려하게 되면, 의미없이 많은 문서에서 사용되는 단어가 가중치가 높아지는 문제점이 발생한다. 따라서 이러한 문제를 방지하기 위해 의미 없는 불용어에 대한 처리가 필요하다. 예를 들어 대부분의 문서에서 사용되는 a, the, is 등의 단어는 낮은 가중치를 주어야 할 것이며, 소수의 문서에서 사용되는 단어는 높은 가중치를 주어야 할 것이다. 역 문서 빈도수는 전체 문서에서 나오는 특정 단어에 대한 빈도수에 대한 가중치를 부여하는 것이다. 역 문서 빈도수의 공식은 다음과 같이 표현될 수 있다.

$$idf(j) = \log\left(1 + \frac{N}{f(j)}\right) \tag{1}$$

위 식에서 N은 문서의 개수이며, $f(j)$ 는 색인어 j가 사용된 문서의 빈도수를 나타낸다. 표 1에서 단어 빈도수 기법을 단순함수로 사용한다면 다음과 같이 가중치를 부여할 수 있다.

$$w(k_{ij}) = f(k_{ij}) \times \log\left(1 + \frac{N}{f(j)}\right) \tag{2}$$

코사인 유사도는 일반적인 유사도 측정 모델로서, 알고리즘이 단순하고 계산이 빠르며, 용어 가중치 기법이 검색 성능을 향상시키고 부분정합으로 질의에 근접한 문서검색이 가능하며 이를 순위화 해준다. 단점으로는 색인어의 상호 독립성 가정이 있다. 식은 두 문서 간 코사인 유사도 계산식이다[10].

$$\begin{aligned}
 & \text{유사도 } (d_x, d_y) = \quad (3) \\
 & \frac{d_x \cdot d_y}{|d_x| \times |d_y|} = \frac{\sum_{a=1}^i w(k_{xa}) \times \sum_{a=1}^i w(k_{ya})}{\sqrt{\sum_{a=1}^i w(k_{xa})^2} \times \sqrt{\sum_{a=1}^i w(k_{ya})^2}}
 \end{aligned}$$

III. 갭 분석 평가모델

본 논문은 기존 교육과정과 능력단위 요소 간 갭 분석의 자동화 평가모델을 제시한다. 평가모델은 각 문서에 대해 형태소 분석과 명사추출을 통해 색인어를 추출하였으며, 단어빈도수-역 빈도수 알고리즘을 이용하여 각 색인어에 가중치를 부여하였다. 또한, 가중치가 부여된 색인어에 대해서 문서 유사도를 효율적으로 측정할 수 있는 코사인 유사도를 이용하여, 능력단위 및 기존 교육과정 간에 유사도를 측정하였으며, 각 유사도에 대해 매핑 테이블을 통해 유사도 평가 결과를 도출하였다. 그림 3에 갭 분석 평가모델에 대해서 나타냈다. 그림 4는 갭 분석 평가모델을 구현하기 위한 컴포넌트 구성도이다.

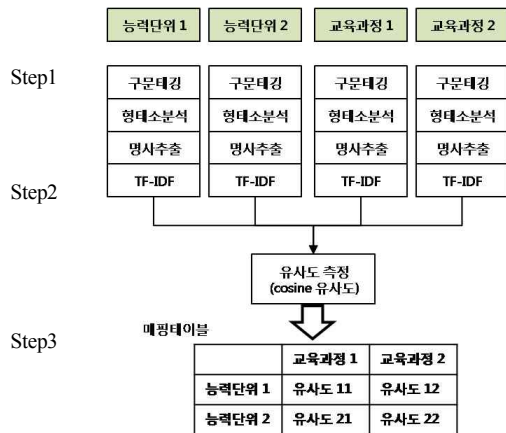


그림 3. 갭 분석 평가모델
Fig. 3. Gap analysis evaluation model.

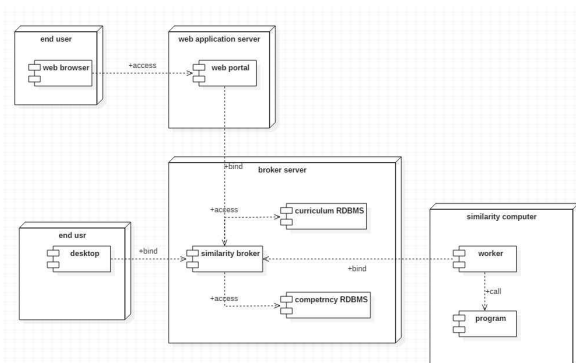


그림 4. 컴포넌트 구성도
Fig. 4. Component configuration.

사용자들이 웹 인터페이스와 또는 데스크 탑 인터페이스를 통해 교과목을 입력하게 되면, 유사도 평가 시스템은 이것을 NCS 능력단위와 비교하여 사용자에게 그 유사도 값을 돌려주게 된다. 그림 5에 갭 분석 평가모델을 위한 전체 소프트웨어 구성에 대해서 나타냈으며, 그림 6에 문서 비교를 위한 주요 함수에 대해서 나타냈다.

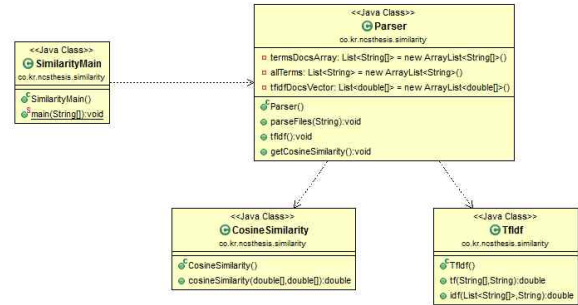


그림 5. 전체 소프트웨어 구성도
Fig. 5. Full software configuration.

```

parseFiles
public void parseFiles(java.lang.String filePath)
    throws java.io.FileNotFoundException,
           java.io.IOException
Method to read files and store in array.
Parameters:
    filePath--:source file path
Throws:
    java.io.FileNotFoundException
    java.io.IOException

tf
public double tf(java.lang.String[] totalTerms,
                java.lang.String termToCheck)
Calculated the tf of term termToCheck
Parameters:
    totalTerms--: Array of all the words under processing document
    termToCheck--: term of which tf is to be calculated.
Returns:
    tf(term frequency) of term termToCheck

idf
public double idf(java.util.List<java.lang.String[]> allTerms,
                java.lang.String termToCheck)
Calculated idf of term termToCheck
Parameters:
    allTerms--: all the terms of all the documents
    termToCheck--
Returns:
    idf(inverse document frequency) score

cosineSimilarity
public double cosineSimilarity(double[] docVector1,
                               double[] docVector2)
Method to calculate cosine similarity between two documents.
Parameters:
    docVector1--: document vector 1 (a)
    docVector2--: document vector 2 (b)
Returns:
    
```

그림 6. 갭 분석 함수
Fig. 6. Gab analysis function.

parseFiles 함수는 기존 교육과정과 능력단위 요소 간 겹 분석을 위해, 폴더에 저장된 모든 능력단위와 기존 교육과정 문서들을 모두 호출하여 파싱한다. 파싱된 단어들은 벡터 공간 모델(vector space model)로 표현하기 위해서, 문자열 벡터 리스트에 저장하였다. 이때, 모든 문서들의 단어들은 중복되는 단어들은 제거하여, 하나의 단어를 하나의 벡터 형식으로 나열하였다. 이것은 각 단어 빈도수-역 빈도수 계산을 위함이다. 각 능력단위와 기존 교육과정 문서들에 대한 파일명은 다음과 같다.

```

competency1.txt      | curriculum1.txt
competency2.txt      | curriculum2.txt
competency3.txt      | curriculum3.txt
...                  | ...
competencyN.txt      | curriculumN.txt
    
```

tf 함수는 각 교육과정 또는 능력단위 요소에 대한 문서에서 단어 빈도를 이용한 가중치를 구하는 기능으로, 다음과 같은 수식을 이용하였다.

$$tf = \frac{\text{단어 빈도수}}{\text{각 문서의 전체 단어 개수}} \quad (4)$$

idf 는 역 문서 빈도를 이용한 가중치를 구하는 기능으로, 다음과 같은 수식을 이용하였다.

$$idf = \log\left(\frac{\text{전체 문서 단어수}}{\text{문서 빈도수}}\right) \quad (5)$$

최종적으로 단어 빈도와 역문서 빈도를 곱하여 단어의 중요도를 나타내는 가중치를 계산하였다. 다음은 코사인 유사도에 대한 주요 프로그래밍 코드이다.

```

dotProduct += docVector1[i] * docVector2[i];
magnitude1 += Math.pow(docVector1[i], 2); //(a^2)
magnitude2 += Math.pow(docVector2[i], 2); //(b^2)
magnitude1 = Math.sqrt(magnitude1); //sqrt(a^2)
magnitude2 = Math.sqrt(magnitude2); //sqrt(b^2)
    
```

IV. 실험 결과

본 연구는 알고리즘을 자바 개발 오픈소스 프로그램인 이클립스에서 자바 언어로 개발하였으며, 테스트는 실제 NCS 학습 모듈과 모 외국어대학교 2015년 1학기 교양영어 강의계획서 내용을 참고하였다. 그림 7에 NCS 학습모듈 검색화면에 대해서 나타냈다[5].

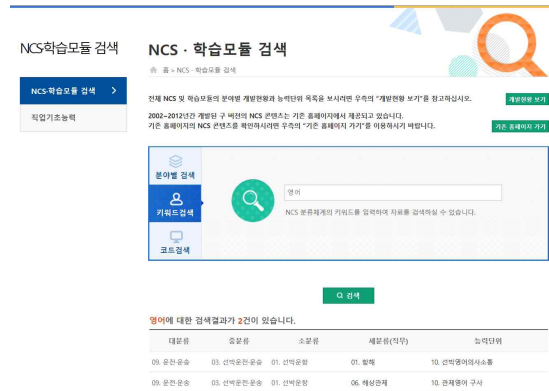


그림 7. NCS 학습모듈 검색화면
Fig. 7. NCS learning module search screen.

NCS 영어 관련 학습모듈을 검색해보면 다음과 같으며, 첫 번째 능력단위 요소 문서인 competency1.txt를 사용하였다.

English communication is even a ship mate, using the anti-feeders, understand meteorological information and correspondence relating to the safe operation of the ship, offense or communicate with coast stations and haegisa with the mission of the multinational crew to use the language knowledge and information to understand English to a standard of English that allows you to perform and the ability to use

관계없는 능력단위 요소 문서를 위해 존 레논의 imagine 일부를 competency2.txt로 사용하였다. 기존 교육과정을 나타내기 위해 모 외국어대학교 2015년 1학기 교양영어 강의계획서를 curriculum1.txt로 사용하였다. 또한, 존 레논의 imagine 일부를 curriculum2.txt로 사용하였다. 다음 내용은 분석된 어휘항목이다.

Course, Summary English, communication, is, even, a, ship, mate, using, the, anti feeders, understand, meteorological, information, and, correspondence, relating, to, safe, operation, of, offense, or, communicate, with, coast, stations, haegisa, mission, multinational, crew, use, language, knowledge, English, standard, that, allows, you, perform, ability, Summary Imagine, theres, no, heaven Its, easy, if, try No, hell, below, us Above, us, only, sky Imagine, all, people living, for, today, Summary This, course, strategically, designed, advancement, students, reading, skills, through, learning, vocabularies, grammar, comprehension, possessions I, wonder, can No, need, greed, hunger A, brotherhood, man imagine, people Sharing, world

다음은 능력단위 요소인 competency1.txt와 competency2.txt 그리고 기존 교육과정인 curriculum1.txt와 curriculum2.txt에 대한 유사도 측정 결과이다.

- 능력단위 1 - 능력단위 1 = 1.0
- 능력단위 1 - 능력단위 2 = 0.0012218162470600004
- 능력단위 1 - 교육과정 1 = 0.03585522348670052
- 능력단위 1 - 교육과정 2 = 0.012443908713423912
- 능력단위 2 - 능력단위 1 = 0.0012218162470600004
- 능력단위 2 - 능력단위 2 = 1.0
- 능력단위 2 - 교육과정 1 = 0.003015365633565146
- 능력단위 3 - 교육과정 2 = 0.1060538808888628
- 교육과정 1 - 능력단위 1 = 0.03585522348670052
- 교육과정 1 - 능력단위 2 = 0.003015365633565146
- 교육과정 1 - 교육과정 1 = 1.0000000000000002
- 교육과정 1 - 교육과정 2 = 0.0062641135618155845
- 교육과정 1 - 능력단위 1 = 0.012443908713423912
- 교육과정 1 - 능력단위 2 = 0.1060538808888628
- 교육과정 1 - 교육과정 1 = 0.0062641135618155845
- 교육과정 1 - 교육과정 2 = 1.0000000000000002

능력단위 1과 2는 능력단위를 기준으로 함으로 생략하고, 교육과정 1과 2 중에서, 우선 교육과정 1에 대해 살펴보면, 능력단위 1과 2에 대한 유사도 값이 각각 0.0358과 0.0124이다. 따라서 교육과정 1은 유사 능력단위로 능력단위 1을 선택할 수 있으며, 능력단위 요소 어느 것과도 유사하지 않은 교육과정 2는 교육과정 편성에서 제외된다. 표 3에 매핑 테이블에 대해서 나타냈다. 그림 8은 개발한 해당 내용을 서비스하는 웹페이지 화면으로서, 기존 교과목 입력 화면이며, 그림 9는 갭 분석 후 결과를 보여주는 화면이다.

표 3. 매핑 테이블
Table 3. Mapping table.

	능력단위 1	능력단위 2	교육과정 1	교육과정 2
능력단위 1	1.0	0.0012	약간유사	유사하지 않음
능력단위 2	0.0012	1.0	유사하지 않음	매우 유사
교육과정 1	약간유사	유사하지 않음	1.0	0.0062
교육과정 2	유사하지 않음	매우 유사	0.0062	1.0000

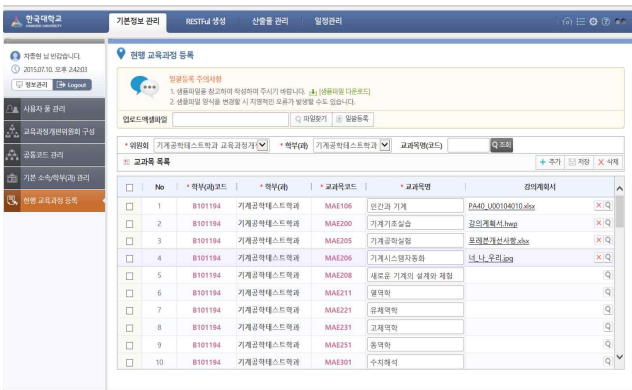


그림 8. 교과목 입력 화면
Fig. 8. Gab analysis service screen.



그림 9. 갭 분석 서비스화면
Fig. 9. Gab analysis service screen.

갭분석 결과 매핑 테이블은 능력단위 요소와 교육과정간의 유사도이며, 갭분석 결과인 표3의 매핑 테이블에서, 교육과정 1과 교육과정 2는 그림 9 갭 분석 서비스 화면에서 제 1영역, 제 2영역으로 열을 채운다. 반면 표3의 매핑 테이블에서 능력단위 1과 능력단위 2는 그림 9 갭 분석 서비스 화면에서 능력단위 요소의 각 해당 항목으로 행을 채운다. 그림 9 갭 분석 서비스 화면에서 만들어진 테이블 각 도메인에는 열과 행을 채운 교육과정과 능력단위 요소의 유사도 척도 값으로 [유사하지 않음], [약간 유사], [매우 유사]가 표시된다. 이러한 유사도는 자동화를 통해 서비스 화면에 표현된다. 이상에서 소개한 유사도 평가모델을 통해, 제 1영역 부터 제 N영역까지 NCS 능력단위와의 유사도 분석을 통해 NCS 교과목으로 선정 여부를 결정한다.

V. 결론

NCS 기반 교육과정 개발을 위해서는 현행 교육과정과 유사 능력단위 요소를 선택하여, 유사도 분석과정인 갭 분석이 필요하다. 기존에 갭 분석은 전문가에 의해 실시되어, 주관적으로 결정되거나, 심리적 요소에 의해 정확성이 결여되고 시간적, 공간적 비효율성이 제기되었다.

본 논문은 기존 교육과정과 능력단위 요소 간 갭 분석의 자동화 평가모델을 제시하였다. 본 논문은 능력단위 및 기존 교육과정 간에 색인어를 추출, TF-IDF를 이용한 가중치 부여, 코사인 유사도를 이용한 유사도를 측정 그리고 매핑 테이블을 통해 유사도 평가 결과를 도출하였다.

본 논문의 평가모델은 본 시스템의 구조적 특징이나 속도 면에서 개선된 알고리즘을 연구해야 한다. 추후 연구로서 빅 데이터를 처리 플랫폼인 하둡(hadoop), 기계학습을 위한 머하우트(mahout) 및 텍스트 기계학습 라이브러리(tm)를 보유한 R 등을 이용하여 다양한 문서 비교 알고리즘을 추가하고자 한다.

참고 문헌

- [1] K. B. Lim, I. J. Ju, and J. Y. Jo, "Pending issue and policy plan of national competency standards," *Korea Research Institute for Vocational Education and Training, The Human Resources Development Review*, Vol. 13, No. 3, pp. 17-39, Sep. 2010.
- [2] D. Y. Park, "Pushing ahead context and project of capability education using national competency standards," *Korea Research Institute for Vocational Education and Training, The Human Resources Development Review*, Vol. 16, No. 3, pp. 52- 71, Sep. 2013.
- [3] D. Y. Kim, J. S. Kim, "Korean national competency standards development trend analysis," *Journal of Korean Technology Education Association*, Vol. 13, No 3, pp. 22-46, Dec. 2013.
- [4] D. K. Kim, S. C. Na, and C. B. Kim, "NCS Learning Module Providing System Using CORS Based on Filter," *Journal of Advanced Navigation Technology*, Vol. 19, No. 2, pp. 161-167, Apr. 2015.
- [5] National Research Foundation of Korea, National competency standards based on NCS curriculum guideline, pp. 7-8, Jan. 2015.
- [6] National Competency Standards, <http://www.ncs.go>
- [7] W. S. Kang, J. S. Lim, J. Y. Kim, S. U. Park, T. H. Lee, and S. B. Hong, "Development of document plagiarism detection algorithm using syntactic analysis method," *Result Report of The Korea Foundation for the Advancement of Science and Creativity*, 2012.
- [8] L. D. Bakerz and A. K. McCallumzy, "Distributional clustering of words for text classification," in *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne: Australia, pp. 96-103, 1998.
- [9] W. M. Song, Y. J. Kim, E. J. Kim, and M. W. Kim, "Development of an efficient key words extraction method using term frequency and word co-occurrence information," in *Proceedings of Korea Intelligent Information System Society Fall Conference*, Seoul: Korea, Vol. 18, No. 2, pp. 193-196, 2008.
- [10] H. S. Kim, S. C. Park, S. H. Kim, "Measurement of Document Similarity using Term/Term-pair Features and Neural Network," *Journal of Korean Institute of Information Scientists and Engineers*, Vol. 31 No. 12, pp. 1660-1671, 2004.



김 대 경 (Dae-kyeong Kim)

1995년 2월 : 인하대학교 국어국문학과(문학사)
 2008년 2월 : 동국대학교 컴퓨터·전자·정보통신공학과(공학석사)
 2012년 2월 : 인천대학교 컴퓨터공학과(공학박사)
 2014년 ~ 현재 : 에듀엔컴(주) 이사, NCS공학연구소 연구소장
 ※ 관심분야 : 소프트웨어공학, 소프트웨어 아키텍처, ICT융합, 지식공학, 인공지능



김 창 복 (Chang-Bok Kim)

1986년 2월 : 단국대학교 전자공학과(공학사)
 1989년 2월 : 단국대학교 전자공학과(공학석사)
 2008년 2월 : 인천대학교 컴퓨터 공학과(공학박사)
 1994년 ~ 현재 : 가천대학교 IT대학 에너지 IT학과 교수
 ※ 관심분야 : 빅 데이터 마이닝, 분산처리시스템, 사물인터넷, 마이크로그리드