
이러닝시스템의 학습 효율성 향상을 위한 색인 메커니즘

김 은 정*

Index Mechanism for advancement learning efficiency of E-learning

Eun-jung Kim*

요 약

오프라인에서는 학습자와 교수자간의 직접적인 학습 내용 전달과 의사소통으로 보다 효율적인 학습이 이루어진다. 이에 가상 학습에서는 특정 학습 영역에 대한 검색 기능과 평가 후 관련 학습 영역으로의 자동 피드백 연결등의 기능으로 이러한 부분을 해결하고 있다. 그러나 검색결과에서 가장 적합한 학습 영역을 선택하고, 현재의 학습을 위해 선행되어야 할 이전 학습 영역을 선택하는 것들이 학습자의 몫이기 때문에 학습 내용에 익숙하지 않은 학습자에겐 오프라인 학습에 비해 많은 어려움이 따른다. 본 논문에서는 학습자가 전체 학습 내용의 흐름과 단원간의 연관성 및 학습 방향을 잡는데 도움을 줄 수 있는 보다 다양한 시각에서의 색인 메커니즘을 제안한다.

ABSTRACT

Offline-based study is proceeding effectively for directly teaching and communication between a learner and a professor. In general virtual education system have been solve this problem using provides search of particular learning domain and auto feedback of relative learning domain after examination. But the learner needs personally select the right document among search result set and interconnected precedence study domain. Therefore, the unskilled learner is difficult of learning progress over against offline-based study. This paper suggests a index mechanism of varied views for helps to learner understand the flow and direction of learning and correlation between units.

키워드

웹기반 교육, 가상 학습 시스템, 원격 교육, 구조적 문서 색인

I. 서 론

기존의 오프라인 형태의 학습 방법은 학습자와 교수자간의 직접적인 학습 내용 전달과 의사소통으로 이루어지기 때문에 학습자의 전반적인 학습 내용 파악이 쉬우며, 부족한 부분에 대한 재학습 영역도 쉽게 판단할 수 있다. 그러나 온라인 가상 학습에서는 교수자가 입력한 학습 내용을 학습자 스스로 학습, 평가, 재학습을 진행하기 때문에 학습의 각 단계별 흐름이나 방향 및 보다 중요한 개념이나 영역 파악, 앞뒤 단원의 연결 그리고 평가에 따른 부족한 영역 파악등에서 많은 어려움이 따른다.

일반적으로 학습이 이루어지는 특정 과목을 살펴보면 각 단원으로 구분되어져 있으며, 각 단원에서의 문단도 그 중요도가 서로 다르다. 또한 앞 단원에서의 개념이나 의미가 다음 단원으로 이어지거나 다음 단원에서 더 상세하게 설명되어지는 경우가 많이 발생한다. 따라서 오프라인 학습 방법에서는 교수자가 이 모든 정보를 직접적인 학습 전달 과정에서 보다 상세하게 학습자에게 전달하므로 학습자의 보다 빠른 이해를 돕는다. 또한 평가 후 틀린 문제에 대한 재학습을 함에 있어서 해당 문제에 대한 설명이나 개념이 특정 단원의 한 문단에 한정될 수도 있고, 여러 단원을 단계별로 다시 학습해야 할 경우도 발생할 수 있다. 이 경우 오프라인 학습에서는 교수자의 설명을 직접 들을 수 있기 때문에 보다 빠르게 재학습 영역을 판단할 수 있다.

본 논문에서는 온라인 가상학습에서 학습자가 스스로 학습과 평가 그리고 재학습을 함에 있어 각 단원별 핵심 내용의 파악과 함께 연관성과 개념 연결 그리고 평가 문제에 관련된 재학습 영역 파악이 보다 효율적으로 이루어질 수 있는 다른 시각에서의 색인 알고리즘을 제시하였다. 이를 위해 학습 내용에서 각 단락별로 교수자가 설명하고자 하는 키워드를 직접 입력하고 이전 영역과 연계되는 경우에는 이전 영역의 구조적 정보를 함께 저장한다. 또한 문서를 색인함에 있어 문서의 전체 내용에서 뿐만 아니라 교수자가 직접 입력한 키워드와 이전 단락과의 연계 정보인 링크를 색인하여 전체적인 학습의 구조 정보를 이루는 새로운 색인 기법을 제시한다.

II. 관련 연구

2.1 온라인 학습시스템에 관한 관련 연구

가상 학습 시스템에서는 효율적인 학습 내용 및 학습과정과 이를 진단하기 위한 평가 방법 및 평가 문제의 객관성 유지관리 그리고 평가후의 재학습의 편리성 등이 핵심이라 할 수 있으며 이에 대한 많은 연구가 있어 왔다 [2-6]. 특히 효율적인 학습 진행을 위한 코스웨어 설계에서는 학습을 위한 문서의 전체 또는 부분 검색 기능과 평가 후 틀린 문제에 대한 재학습의 기능도 제공하고 있다.

대부분의 기존 시스템에서 문서의 색인 구조에 대한 세부적인 기법은 설명되어지지 않고 있다. 하지만 기존 시스템의 결과 화면과 설명을 요약하면, 첫째 문서의 내용에서 색인어를 추출하여 학습 내용의 전체 또는 부분 검색을 제공한다. 둘째 평가 문제를 등록할 때 교수자가 해당 문제와 관련된 학습 영역을 등록하여, 학습자가 틀린 문제와 관련된 학습 영역으로 자동 피드백을 가능하게 한다. 이러한 기존 시스템에서는 특정 학습 영역의 검색 결과에서 가장 관련된 학습 내용의 선택은 학습자 스스로 판단해야 한다. 또한 정확한 학습 영역을 선택했을 경우에도 해당 학습 내용에 대한 사전 지식이 없는 학습자의 경우에는 현재의 학습 내용을 위해서 선행되어야 할 이전 학습 영역을 스스로 판단하기란 쉽지가 않다. 그리고 평가에 따른 재학습의 경우에도 평가문제와 관련된 한 영역으로만 피드백되기 때문에 관련된 이전 학습 영역을 선택하는 것은 학습자의 몫이므로 이 또한 많은 어려움이 따른다.

본 논문에서는 학습 내용을 구조적 문서로 작성하여 관리한다. 교수자가 학습 내용을 입력함에 있어서 각 단락별로 중요한 키워드들을 입력하여 문서의 색인 과정에서 문서의 전체 내용에서 색인어를 추출하기 보다는 최소 단위의 단락 내용에서 색인어를 추출하며, 또한 교수자의 입력 키워드 정보에서도 색인어를 추출한다. 그리고 현재의 단락과 연계되는 개념 및 이론이나 반드시 선행 학습해야 할 이전 단락의 링크 정보를 입력하여 각 단락의 학습 연계성을 의미하는 링크 정보도 색인한다.

2.2 일반 벡터 검색 모델

문서나 하이퍼텍스트 노드안에 있는 개념들이나 또는 각 단일 용어들이 갖는 상대적인 중요성을 측정하기 위해 일반 벡터 공간 모델에서는 다음의 작업을 수행한

다. 문서 i 에서 용어 j 에 대한 가중치를 계산하는 방법은 식 (1)과 같다.

$$w_{ij} = tf_{ij} \cdot idf_j = tf_{ij} \cdot \log\left(\frac{n}{df_j}\right) \quad (1)$$

식 (1)에서, w_{ij} 는 문서 i 에서 용어 j 의 가중치이며, tf_{ij} 는 문서 i 에서 용어 j 의 빈도수, idf_j 는 용어 j 에 대한 역문헌 빈도수이다. n 은 전체 문헌수, df_j 는 용어 j 를 가지는 문서 수이다. 식 (1)의 결과로 나온 검색 결과 집합에서, 질의어 Q 에 대하여 문서의 순위를 계산하기 위하여 각 문서의 RSV (Retrival Status Value)의 계산은 식 (2)와 같다.

$$RSV(D_i, Q) = \sum_{j=1}^q w_{ij} \cdot w_{qj} \text{ for } i = 1, 2, \dots, n \quad (2)$$

각 문서 D_i 는 질의어 Q 에 대하여, 질의어 Q 에 있는 모든 용어 j 의 문서 가중치 w_{ij} 와 질의어 가중치 w_{qj} 의 곱에 대한 누적 합산의 결과로 RSV 를 갖는다. w_{ij} 는 식 (1)에서 정의한 내용이고, w_{qj} 는 질의어 q 에서 용어 j 가 갖는 가중치이다. 이렇게 얻어진 RSV 값에 따라 각 문서는 질의어 Q 에 대해 순위를 갖게 된다.

III. 학습내용을 위한 XML문서와 색인 구조

3.1 학습 내용 DTD

과목별 학습 내용을 위한 XML DTD는 그림 1과 같다. 본 논문에서는 소단원(unit)이나 또는 각 단락(paragraph)마다 교수자가 해당 내용을 대표할 수 있는 키워드(keyword)와 연계된 앞 단원의 정보에 대한 링크(pre_link)를 입력한다. 이는 검색 결과 집합에서의 문서의 우선순위를 정할 때와 특정 학습 영역의 검색이나 평가 후 특정 문제의 재학습을 위해서 1장부터 마지막장까지 관련된 단원을 연계해서 보여 줄 때 사용한다.

```
<!ELEMENT course (subject)>
<!ELEMENT subject (sub_title, part+ | chapter+) >
<!ELEMENT sub_title (#PCDATA)>
<!ELEMENT part (part_title, chapter+) >
<!ELEMENT part_title (#PCDATA)>
<!ELEMENT chapter (chap_title, unit+ | paragraph+) >
<!ELEMENT chap_title (#PCDATA)>
<!ELEMENT unit (unit_title, paragraph+, keyword*, pre_link*) >
<!ELEMENT unit_title (#PCDATA)>
<!ELEMENT paragraph (content, keyword*, pre_link*) >
<!ELEMENT content (#PCDATA) >
<!ELEMENT keyword (#PCDATA) >
<!ELEMENT pre_link (#PCDATA) >
```

그림 1. 과목별 XML DTD인 course.dtd
Fig. 1. course.dtd of XML DTD by subjects

3.2 XML 문서 내용 및 구조

그림 2는 그림 1의 과목별 XML DTD를 기반으로 특정 과목의 학습내용을 입력한 예이다. 그림 2에서 '9.2 클래스'의 첫 번째 단락에 키워드와 링크가 입력되어 있다. 링크의 "prg1.xml#part/chapter_7/unit/paragraph_3"은 prg1.xml 문서, 7장, 1단원, 3번째 단락으로의 링크의 의미를 가진다. 이 레이터는 문서의 색인 과정에서 문서 구조 정보에 부여되는 각 엘리먼트의 색인 ID에 의해 prg1.xml#(0,1,7,1,3,0)로 저장된다.

```
<?xml version="1.0" encoding="euc-kr"?>
<!DOCTYPE course SYSTEM "course.dtd">
<course> <subject>
<sub_title> 프로그래밍 언어 일반 </sub_title>
<part> <part_title> 프로그래밍 언어 </part_title>
  <chapter>
    <chap_title> 1. 프로그래밍 언어의 개요 </chap_title>
    <unit> <unit_title> 1.1. 프로그래밍 언어의 정의 </unit_title>
      <paragraph> <content>
        프로그래밍언어란 주어진 문제의 해결을 위한 절차를 표현한 알고리즘과 이 알고리즘의 실행시 필요한 레이터를 설명하는 표현이라...
      </content>
      <keyword> 프로그래밍언어 정의 </keyword>
    </paragraph>
  </unit> .....
</chapter> .....
<chapter>
  <chap_title> 9. 객체지향 프로그래밍 </chap_title>
```

```

.....
<unit> <unit_title> 9.2. 클래스 </unit_title>
<paragraph> <content>
객체를 사용하기 위해 먼저 클래스를 정의해야 한다.
클래스 내부변수와 내부변수를 조작할 함수를 .....
</content>
<keyword>클래스정의,클래스의미,클래스개념
</keyword>
<pre_link> prg1.xml#part/chapter_7/unit/paragraph_3
</pre_link>
</paragraph>
</unit>
</chapter>
</part>
</subject>
</course>
    
```

그림 2. 과목별 XML 문서인 prg2.xml
Fig. 2. pro2.xml of XML document by subjects

그림 3은 그림 2의 문서 구조를 그림으로 표현한 것이다. 타원은 엘리먼트를 나타내고, 직사각형은 실질적으로 들어가는 문서의 내용이다. 여기서 최상위 엘리먼트를 인덱스 [0]으로 표현하고, 그 자식 엘리먼트를 순서대로 [0,0], [0,1], [0,2],...로 나타낼 수 있다. 계속해서 [0,1]의 자식 엘리먼트는 [0,1,0], [0,1,1],...순으로 나타낸다. 그림 3에서 9장의 첫번째 단락([0,1,9,2,1,2])의 엘리먼트 pre_link의 입력 내용은 이러한 구조상에서 이루어진 인덱스 번호가 주어진다.

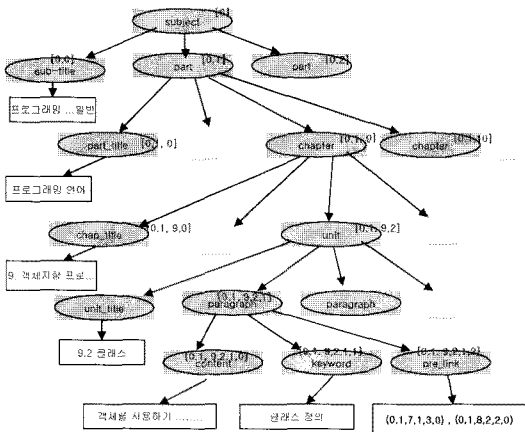


그림 3. XML 문서의 구조
Fig. 3. Structure of XML document

3.3 색인 구조

본 논문의 문서 구조를 이용한 학습 시스템을 위한 색인 구조는 역화일[1]과 동등한 구조를 가진다. 그림 4는 prg2.xml 문서에 대한 구조 정보 파일 구조이다. 색인ID는 문서내 각 엘리먼트를 구별하기 위한 유일한 식별 번호이며, 엘리먼트는 해당 ID를 갖는 문서내 엘리먼트 위치를 의미한다. 이전링크는 <paragraph> 단위에서 관련 있는 이전 단락의 주소를 가진 경우, 해당 주소에 대한 정보를 저장한다. 그림 4에서 (0,1,7,1,3, 0)은 같은 문서의 색인ID이고, prg1.xml#(0,1,5,2,4)는 prg1.xml문서의 특정 엘리먼트를 지시한다.

색인 ID	엘리먼트	이전링크
(0)	subject	
(0,1)	subject/part_1	
(0,2)	subject/part_2	
(0,1,0)	subject/part_1/part_title	
....		
(0,1,9)	subject/part_1/chapter_9	
....		
(0,1,9,2)	subject/part_1/chapter_9/unit_2	
(0,1,9,2,1)	subject/part_1/chapter_9/unit_2 /paragraph_1	(0,1,7,1,3,0)
(0,1,9,2,1,0)	subject/part_1/chapter_9/unit_2 /paragraph_1/content	
....		prg1.xml# (0,1,5,2,4)

그림 4. prg2.xml 구조 정보 파일
Fig. 4. Structure information file of prg2.xml

다음으로 색인 화일에는 문서 단락을 대표하는 색인어들을 B-Tree 형식으로 저장하고 색언어가 가리키는 포스팅 화일에는 실제 해당 색언어가 나타난 문서의 단락에 대한 정보를 가지고 있다. 그림 5는 이러한 색인 화일과 포스팅 화일에 대한 구조이다. 포스팅 파일의 type 'c'는 문서의 내용에서, 'k'는 교수자가 입력한 <keyword>에서 추출한 색언어임을 의미한다. 이는 검색 결과 집합에 대한 우선순위를 부여할 때 사용된다.

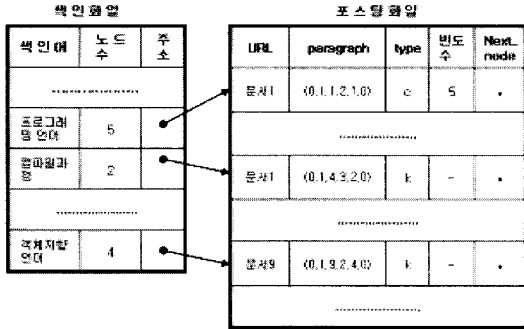


그림 5. 색인 구조
Fig. 5. Index structure

3.4 색인 과정 및 알고리즘

그림 6에서 하나의 문서 내용을 색인하기 위해 3단계의 과정을 수행한다.

```

indexing(xml_document) {
  for each document d /* 문서전체 각 엘리먼트의 구조정보 저장*/
    문서에서 엘리먼트를 선택;
    엘리먼트에 index_ID를 부여;
    엘리먼트 정보를 구조정보 파일에 추가;
    if (엘리먼트가 <pre_link>)
      <pre_link>의 내용을 구조 정보 파일에 추가;
  }
  for each document d /* 엘리먼트<keyword> 내용 색인 저장*/
    엘리먼트 <keyword>에서 색인어를 선택;
    if ( 색인어가 색인화일에 등록되지 않았는가?)
      색인어를 색인화일에 등록;
      색인어 정보(색인어, 'k')를 포스팅 화일에 추가;
  }
  for each document d { /* 문서내용 색인 저장*/
    문서 내용 <paragraph>단위에서 색인어를 선택;
    if ( 색인어가 색인화일에 등록되지 않았는가?)
      색인어를 색인화일에 등록;
      색인어 정보(색인어, 'c', 빈도)를 포스팅 화일에 추가;
  }
}
    
```

그림 6. 색인 알고리즘
Fig. 6. Index algorithm

첫째, 문서의 구조 정보를 저장하기 위해 DTD 파일을 참조하여 첫 번째 엘리먼트부터 색인 ID를 부여하고 엘리먼트 위치 정보를 저장한다. 그리고 엘리먼트 <unit>, <paragraph> 단위에서는 엘리먼트 <pre_link>에 대한 정보를 이용하여 연관된 이전 학습에 대한 링크 정보를 저장하여 전체적인 문서의 관련 정보 흐름을 파악할 수 있게 한다. 이 과정을 문서 내 모든 엘리먼트에 대해 수행함으로써 하나의 문서에 대한 구조 정보를 저장할 수 있다. 둘째, 문서내 엘리먼트 <keyword>에 대한 색인 과정을 먼저 수행한다. 포스팅 파일의 url은 xml 문서의 위치정보를, paragraph은 문서의 구조 색인 ID, type은 색인어가 엘리먼트 <keyword>에 있는 내용을 의미하는 'k'를 저장한다. 마지막으로 문서의 내용에서 <paragraph> 단위로 색인어를 찾는 과정을 수행한다. 포스팅 파일의 type은 색인어가 문서 내용에 있음을 의미하는 'c'를 저장하고, 빈도수에는 해당 단락에서 색인어가 나타난 회수를 저장한다. 계속해서 단어가 나타날 때마다 이 과정을 반복한다.

IV. 실험 결과 및 분석

이전의 연구[5,6]에서 웹기반 기사시험 학습 시스템을 설계한 바 있다. 본 논문에서 제시한 색인 기법의 효율성을 실험하기 위해 이전 학습 시스템을 이용하여 기존 색인 방법과 본 논문의 색인 방법을 실험하여 그 결과를 비교 분석하였다.

이전 시스템에서는 학습 내용을 HTML 문서로 저장하여 운영하였으며, 관련 연구에 있는 대부분의 기존의 가상 학습 시스템과 마찬가지로 문서의 전체 검색과 특정 영역의 검색을 제공하고 있다. 문서의 특정 영역으로의 링크는 HTML 태그에서 <A>태그의 name 속성을 이용하였다. 학습자의 특정 검색어에 대한 검색 결과 집합에서 문서내 검색어의 빈도수를 이용하여 문서의 우선순위를 부여하였다. 이전 학습 시스템에서의 검색어에 대한 결과 화면은 그림 7과 같다.

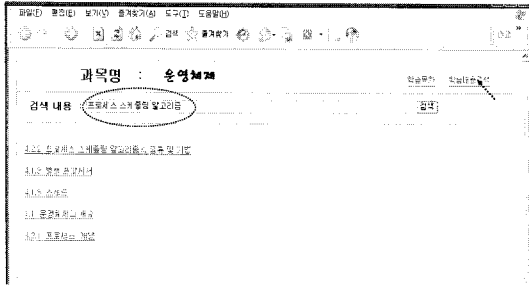


그림 7. HTML 문서의 검색 결과 화면1
Fig. 7. Search result screen 1 of HTML document

그림 7의 우선 순위 결과는 문서내 '프로세스', '스케줄링', '알고리즘' 용어의 빈도수가 큰 순서로 부여된다. 첫번째 문서는 '프로세스 스케줄링 알고리즘'에 대한 학습 내용이 있는 문서이며, 교수자의 주관적인 의도에서도 해당 검색어에 추천을 하고자 하는 문서이다. 그러나 모든 검색어에 대한 검색 결과에서 항상 가장 적합한 문서가 높은 우선 순위를 가지는 것은 아니며, 또한 문서의 우선순위가 학습을 원하는 내용의 우선순위가 아닐 수 있기 때문에 학습자는 검색 결과에서 스스로 관련 내용을 찾아서 학습을 수행해야 한다. 그리고 그림 8에서 가장 적합한 문서로 학습하는 경우에도 관련된 선행 학습을 위해서는 문서내 특정 링크를 선택하거나, 다시 검색어를 이용해서 선행 학습 영역을 판단해야 한다.

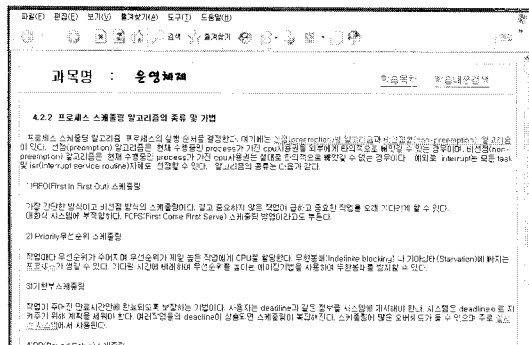


그림 8. HTML 문서의 검색 결과 화면2
Fig. 8. Search result screen 2 of HTML document

다음으로 본 논문에서는 오프라인에서 교수자가 직접 전달하는 학습의 방향과 진도대로 학습자를 이끌 수

있는 색인 기법을 제시하였다. 그림 9의 결과에서 문서7은 '프로세스 스케줄링 알고리즘'이라는 색인어로 교수자의 키워드에 의해 색인되어 검색어에 대한 검색 결과 집합에서 우선하여 나타난다. 그리고 그림 10은 검색어 'kernel'에 대하여 교수자가 입력한 키워드가 없는 경우이며 문서의 내용에 나타난 용어의 빈도수에 의해 우선 순위가 결정되어 결과가 나타난다.

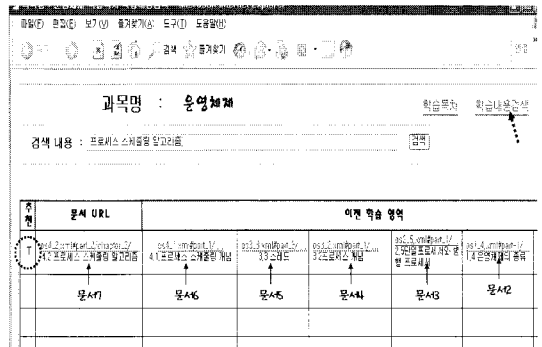


그림 9. XML문서의 검색 결과 화면1
Fig. 9. Search result screen 1 of XML document

그림 9에서 첫번째 필드 [추천]의 'T'는 문서7이 검색어 그대로 교수자가 입력한 키워드에서 찾은 즉, 교수자가 추천한 문서임을 의미한다. 또한 학습 영역의 링크 정보를 이용하여 문서7->문서6->문서5->문서4->문서3->문서2의 순서로 선행되어져야 할 이전 학습 영역의 정보를 함께 제공함으로써 학습자가 backtracking 기법으로 학습을 효율적으로 행할 수 있다.

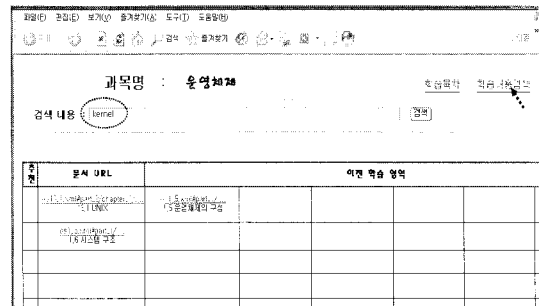


그림 10. XML문서의 검색 결과 화면2
Fig. 10. Search result screen 2 of XML document

V. 결론

본 논문에서는 온라인 가상 학습에서 학습 내용에 익숙치 않은 학습자들이 전체 학습 내용의 흐름과 단원간의 연계성 및 학습 방향을 파악하여 학습 진행을 효율적으로 진행하는데 도움을 주고자 보다 다양한 시각에서의 색인 알고리즘을 제안하였다. 이는 학습 내용 검색과 선행 학습 영역의 선택에 도움을 줄 수 있으며, 평가 문제에 따른 재학습에서도 각 단원별 선행 학습 정보를 쉽게 이용할 수 있다.

참고문헌

- [1] William B. Frakes and Ricardo Baeza-Yates, "Information Retrieval Data Structures & Algorithms", Prentices Hall, 1992
- [2] 정용기, 최은만, "웹 기반 학습평가 자동화 시스템의 설계 및 구현", 한국정보처리학회 논문지, pp.289-296, 2002년 4월.
- [3] 고일석외3, "웹 기반 가상학습 시스템의 설계 및 구현", 한국정보처리학회 논문지, pp.631 - 638, 2002년 12월.
- [4] 김경아, 최은만, "웹기반 교육에서의 자동 문제출제 시스템", 한국정보처리학회 논문지, pp.301-310, 2002년 9월.
- [5] 류희열, 김은정, "기사 시험을 위한 웹기반 학습 시스템의 설계", 한국정보과학회 봄학술발표논문집, 2004년 4월.
- [6] 김은정, "웹기반 학습 시스템의 평가 문제에 대한 출제 방법 및 난이도 재조정에 대한 연구", 정보처리학회 논문지, 2005.

저자소개



김은정(Eun-Jung Kim)

2001년 경상대학교(공학박사)
1990~1993년 LG전자 멀티미디어 연구소 연구원

2001~2005년 부산외국어대학교 전자컴퓨터공학부 초빙교수, 전임강사(강의전담)

2006~2007년 부산대학교 시간강사

2008~현재 부산가톨릭대학교 컴퓨터공학과 전임강사

※관심분야: 정보검색, 원격교육, 웹기반 가상학습