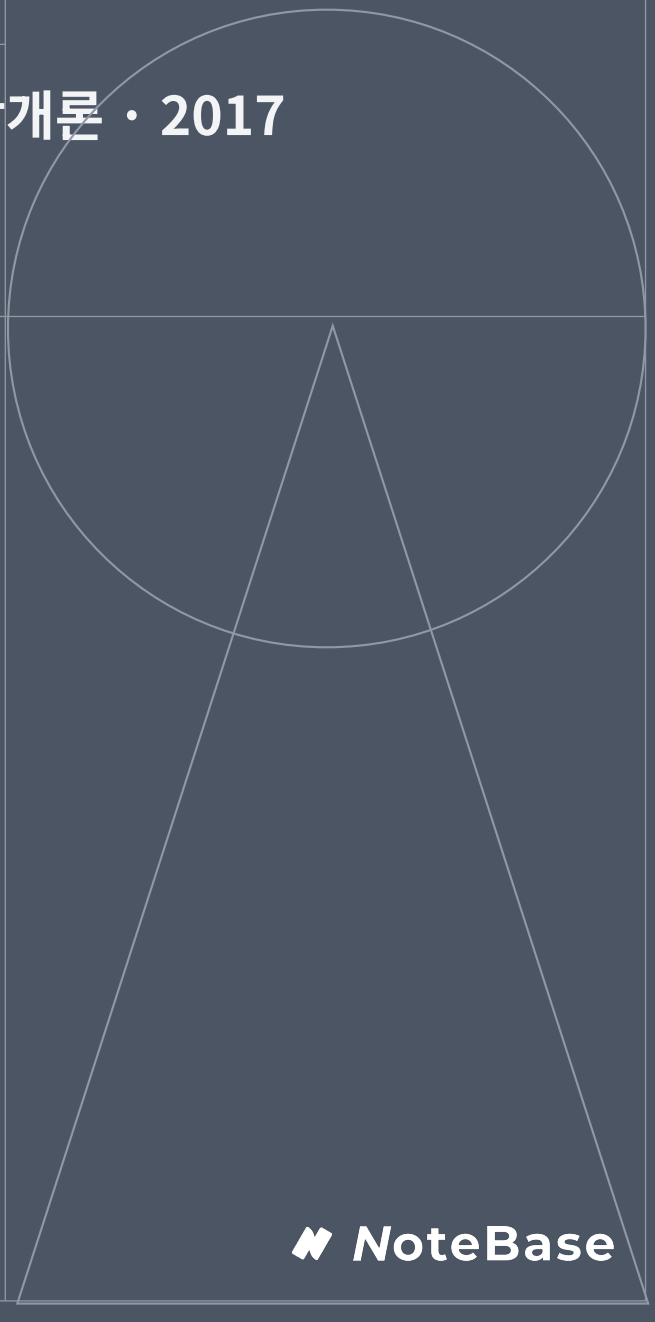
기출OX 해설집

공무원/국회직/9급 · 정보학개론 · 2017



NoteBase

기출OX 해설집

공무원/국회직/9급 · 정보학개론 · 2017

공무원/국회직/9급 · 정보학개론 · 2017

1. **정보표현과 검색에 사용하는 언어 또는 도구에 대한 설명으로 옳지 않은 것은?**

① 시소러스 - 용어들의 통제어휘이며 후조합언어이기 때문에 이용자가 표현과 검색 시점에서 용어를 조합할 수 있다.

② 분류표 - 주로 전조합을 위해 만들어진 수문자 형태의 통제어휘이다.

③ 택소노미 - 용어들 간의 상호참조 기능을 제공하며 계층적 관계를 표현해 준다.

④ 폭소노미 - 이용자들에 의해서 자유롭게 선택된 키워드들을 이용해서 이루어지는 협업적 분류이다.

⑤ 온톨로지 - 개념 간의 관계를 정의하고 개념을 추론하기 위한 논리적 규칙을 명시하는 서술의 집합이다.

정답 3번

출제 포인트

시소러스, 분류표, 택소노미, 폭소노미, 온톨로지가 용어를 통제하고 개념 관계를 표현하는 방식을 구별하는 문제이다.

핵심 개념

시소러스: 색인어의 동의·계층·연관 관계를 제시하는 통제어휘이며, 검색 시 용어를 조합하는 후조합 방식에 활용된다.
분류표: 지식을 계층적으로 배열하고 분류기호로 나타내는 통제어휘이며, 미리 결합된 분류 표현을 사용하는 전조합 방식이 중심이다.
택소노미: 개념이나 대상을 상위·하위 범주로 조직하는 계층적 분류체계이다.
폭소노미: 이용자가 자유롭게 붙인 태그를 모아 형성하는 협업적 분류이다.
온톨로지: 특정 영역의 개념, 속성, 관계와 규칙을 형식적으로 명시하여 추론이 가능하게 하는 지식 표현 체계이다.

NoteBase

선지별 해설

① O — 시소러스는 통제어휘이고, 개별 용어를 검색 시점에 조합할 수 있는 후조합 언어로 활용된다.
② O — 분류표는 지식을 계층적으로 배열한 통제어휘이며, 분류기호를 이용해 주제를 미리 결합하여 표현하는 전조합 방식이 중심이다.
③ X — 틀린 표현은 '용어들 간의 상호참조 기능을 제공하며'이다. 상호참조를 통해 동의·계층·연관 관계를 연결하는 기능은 주로 시소러스의 특징이다. 택소노미는 개념이나 대상을 상위 범주와 하위 범주로 조직하여 계층적 관계를 표현하는 분류체계라고 고쳐야 한다.
④ O — 폭소노미는 이용자가 통제되지 않은 키워드나 태그를 자유롭게 선택하고 공유하여 형성하는 협업적 분류이다.
⑤ O — 온톨로지는 개념과 개념 사이의 관계를 명시하고, 그 관계와 논리 규칙을 바탕으로 새로운 지식을 추론할 수 있게 한다.

오답 포인트

③은 택소노미의 '계층적 관계'라는 올바른 특징에 시소러스의 핵심 기능인 '용어 간 상호참조'를 섞었다. 두 용어 모두 계층 관계를 다룰 수 있지만, 관계어를 연결하여 검색을 돕는 상호참조 체계는 시소러스에 더 직접적으로 대응한다.

다시 나오면 이렇게 판단

용어 간 동의·계층·연관 관계와 상호참조는 시소러스, 상위·하위 범주의 계층적 조직은 택소노미, 이용자 자유 태그는 폭소노미, 개념 관계와 규칙에 따른 추론은 온톨로지로 구별한다.

암기 포인트

시소러스는 용어 관계와 참조, 택소노미는 계층 분류, 폭소노미는 자유 태그, 온톨로지는 관계와 추론이다.

NoteBase

2. 정보요구란 '어떠한 문제상황에서 이용자가 그 문제를 해결하기 위하여 정보가 필요함을 인지하게 되는 과정'이라 정의하고, 이러한 '지식의 이상상태에 처한 이용자가 그 부족함을 해결하기 위해 정보를 찾아나서게 된다'고 본 학자는?

① R. Taylor
② N. Belkin
③ M. Voigt
④ H. Borko
⑤ V. Bush

정답 2번

출제 포인트

정보요구를 문제상황에서 필요를 인식하는 과정으로 보고, 이용자의 지식이 비정상적·불완전한 상태에 놓일 때 정보탐색이 시작된다고 설명한 학자를 찾는 문제이다.

핵심 정리

N. Belkin: 이용자가 자신의 지식에 이상 상태(ASK, Anomalous State of Knowledge)가 있음을 인식하면 그 부족함을 해소하기 위해 정보를 탐색한다고 본다.
R. Taylor: 이용자의 정보요구가 내면적 요구에서 의식적 요구, 형식화된 질문, 정보시스템에 제시하는 타협된 질문으로 구체화되는 단계를 설명한다.
M. Voigt: 과학자가 정보를 찾는 접근을 최신정보 파악, 일상적 정보 탐색, 특정 문제에 대한 포괄적 탐색 등으로 구분한다.
H. Borko: 정보과학을 정보의 속성과 행태, 정보 흐름을 지배하는 힘, 정보 처리 수단을 연구하는 학문으로 정의한다.
V. Bush: 인간의 연상 작용을 본뜬 정보 저장·검색 장치인 메멕스(Memex)를 제안한다.

NoteBase

선지별 해설

① X — R. Taylor라는 연결이 틀리다. Taylor는 정보요구가 내면적 요구에서 실제 질문으로 구체화되는 단계를 설명한 학자이며, 제시문의 이상 지식 상태 모형은 N. Belkin과 연결된다.
② O — N. Belkin은 이용자가 문제상황에서 자신의 지식이 불완전하거나 비정상적인 상태임을 인식하고, 이를 해결하기 위해 정보를 찾는다고 보는 ASK 모형을 제시한다.
③ X — M. Voigt라는 연결이 틀리다. Voigt는 과학자의 정보탐색 접근 유형을 설명한 학자이며, 제시문의 정의는 N. Belkin의 ASK 모형에 해당한다.
④ X — H. Borko라는 연결이 틀리다. Borko는 정보의 속성·행태와 정보의 흐름 및 처리 수단을 연구하는 학문으로 정보과학을 정의했으며, 제시문의 학자는 N. Belkin이다.
⑤ X — V. Bush라는 연결이 틀리다. Bush는 연상 색인을 활용하는 가상 정보 장치 메멕스를 제안했으며, 제시문의 정보요구 관점은 N. Belkin의 것이다.

출제자가 노린 함정

선지의 인물들은 모두 정보학의 주요 학자이므로 이름만 익숙하면 혼동하기 쉽다. 제시문의 결정적 표현은 '지식의 이상상태'와 그 부족함을 해결하기 위한 정보탐색이며, 이는 Belkin의 ASK를 가리킨다.

다시 나오면 이렇게 판단

'문제상황—지식의 이상상태—정보탐색'이 함께 제시되면 Belkin의 ASK로 판단한다. 정보요구가 질문으로 구체화되는 단계는 Taylor, 정보과학의 학문적 정의는 Borko, 메멕스는 Bush로 구분한다.

암기 포인트

Belkin = ASK = 지식의 이상상태를 해소하기 위한 정보탐색이다.

NoteBase

3. <보기>의 괄호 안에 들어갈 용어는?

<보 기>

어떤 연구자의 논문들 중 ◯개의 논문이 최소 ◯회 이상씩 인용되었을 경우 그 연구자의 ( )는 ◯이다.

① Eigenfactor
② Impact factor
③ Scimago
④ H-index
⑤ SNIP

정답 4번

출제 포인트

이 문제는 연구자의 논문 수와 피인용 횟수를 함께 반영하는 H-index의 정의를 다른 학술지 영향력 지표와 구별하는 문제이다.

핵심 개념

H-index: 어떤 연구자의 논문 가운데 h편이 각각 최소 h회 이상 인용되었을 때의 최대 h값으로, 연구자의 연구 생산성과 인용 영향력을 함께 나타내는 지표이다.
Eigenfactor: 학술지 인용망에서 인용의 출처와 영향력을 가중하여 학술지의 전체 영향력을 평가하는 지표이다.
Impact factor: 특정 기간에 학술지에 게재된 논문이 받은 평균 인용 횟수를 바탕으로 산출하는 학술지 지표이다.
SCImago Journal Rank(SJR): 인용한 학술지의 위신까지 반영하여 학술지의 영향력을 평가하는 지표이다.
SNIP: 학문 분야별 인용 관행의 차이를 보정하여 학술지의 논문당 인용 영향력을 평가하는 지표이다.

NoteBase

선지별 해설

① X — Eigenfactor는 연구자의 논문 수와 최소 피인용 횟수가 같은 값을 나타내는 지표가 아니다. 틀린 용어는 ‘Eigenfactor’이며, 올바르게는 학술지 인용망을 바탕으로 학술지의 전체 영향력을 평가하는 지표이다. 빈칸에는 H-index가 들어간다.
② X — Impact factor는 개별 연구자의 h편 논문이 각각 h회 이상 인용되었는지를 나타내지 않는다. 틀린 용어는 ‘Impact factor’이며, 올바르게는 학술지 논문의 평균 피인용 수준을 나타내는 지표이다. 빈칸에는 H-index가 들어간다.
③ X — Scimago는 제시된 연구자 단위 지표가 아니다. 틀린 용어는 ‘Scimago’이며, 올바르게는 SCImago Journal Rank(SJR)가 인용한 학술지의 위신을 반영해 학술지의 영향력을 평가하는 지표이다. 빈칸에는 H-index가 들어간다.
④ O — H-index가 h라는 것은 그 연구자의 논문 중 h편이 각각 최소 h회 이상 인용되었다는 뜻이다. 제시문의 두 빈칸과 지표 값에 같은 수가 들어가는 정의와 정확히 일치한다.
⑤ X — SNIP는 연구자의 논문 수와 최소 인용 횟수를 짝지어 나타내는 지표가 아니다. 틀린 용어는 ‘SNIP’이며, 올바르게는 분야별 인용 관행을 보정한 학술지의 논문당 인용 영향력 지표이다. 빈칸에는 H-index가 들어간다.

오답 포인트

선지의 용어들이 모두 인용을 활용하므로 같은 종류의 지표처럼 보이는 것이 함정이다. 제시문은 ‘h편의 논문이 각각 최소 h회 이상 인용’이라는 연구자 단위의 정의를 제시하므로, 학술지 단위 지표인 Eigenfactor, Impact factor, SJR, SNIP와 구별해야 한다.

다시 나오면 이렇게 판단

논문 편수와 각 논문의 최소 피인용 횟수에 같은 수가 반복되면 H-index로 판단한다. 먼저 평가 대상이 연구자인지 학술지인지 구분하면 유사한 인용 지표를 빠르게 가를 수 있다.

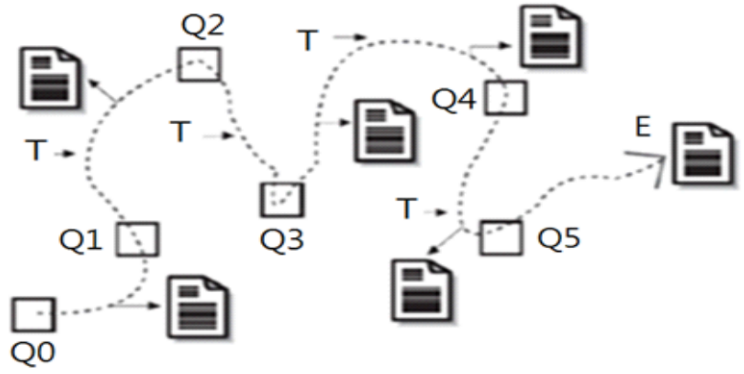
암기 포인트

H-index h = 최소 h회 인용된 논문이 h편이다.

NoteBase

4. <보기>와 같이 검색을 전개하는 탐색기법은?

<보 기>



① Snowballing
② Berrypicking
③ Citation Pearl Growing
④ Building Block
⑤ Lowest Posting First

정답 2번

출제 포인트

이 문제는 검색 과정에서 생각(T)과 새로 얻은 문헌·정보에 따라 질의가 Q0에서 Q5까지 계속 변하고, 필요한 정보를 조금씩 얻은 뒤 종료(E)하는 도식을 보고 탐색기법을 식별하는 문제이다. 이는 검색자가 여러 정보원에서 유용한 정보 조각을 골라 가며 질의를 발전시키는 Berrypicking이다.

NoteBase

선지별 해설

① X — Snowballing은 이미 찾은 핵심 문헌의 참고문헌이나 인용 문헌을 따라 관련 문헌을 연쇄적으로 확장하는 방법이다. 도식처럼 생각과 검색 결과에 따라 질의 자체를 Q0부터 Q5까지 거듭 바꾸는 기법은 Snowballing이 아니라 Berrypicking이다.
② O — Berrypicking은 검색 과정에서 얻은 단서와 정보에 따라 질의를 계속 수정하고, 여러 문헌과 정보원에서 필요한 부분을 조금씩 선택하여 탐색을 전개한다. Q0→Q1→⋯→Q5의 질의 변화와 중간중간의 문헌 획득이 이 특징을 나타낸다.
③ X — Citation Pearl Growing은 매우 적합한 문헌 한 건을 '진주'로 삼아 그 문헌의 주제어·색인어 또는 인용 관계를 이용해 관련 문헌을 늘려 가는 방법이다. 제시된 도식의 핵심은 한 핵심 문헌에서 확장하는 것이 아니라 검색 도중 질의를 계속 변화시키는 Berrypicking이다.
④ X — Building Block은 검색 주제를 여러 개념 블록으로 나누고, 같은 개념의 동의어는 OR로 묶으며 서로 다른 개념 블록은 AND로 결합하는 방법이다. 도식은 개념 블록의 논리 결합이 아니라 질의의 연속적인 변화를 보여 주므로 Berrypicking에 해당한다.
⑤ X — Lowest Posting First는 검색어 또는 검색 집합 가운데 포스팅 수가 가장 적은 것부터 처리하여 검색 범위를 효율적으로 줄이는 방식이다. 질의 변형과 정보 조각의 선택을 반복하는 제시 도식은 Lowest Posting First가 아니라 Berrypicking이다.

오답 포인트

문헌을 연쇄적으로 찾아간다는 겉모습만 보면 Snowballing이나 Citation Pearl Growing과 혼동하기 쉽다. 그러나 이 도식의 결정적 단서는 Q0, Q1, Q2, Q3, Q4, Q5처럼 검색 과정에서 질의가 계속 달라지고, 각 단계에서 얻은 문헌과 생각이 다음 검색을 바꾼다는 점이다.

다시 나오면 이렇게 판단

질의가 검색 도중 계속 진화하고 여러 정보원에서 필요한 조각을 골라 모으면 Berrypicking으로 판단한다. 한 핵심 문헌의 인용 관계를 따라가면 Snowballing, 한 적합 문헌의 색인어나 인용 정보를 출발점으로 확장하면 Citation Pearl Growing으로 구분한다.

암기 포인트

Berrypicking은 '질의를 바꾸며 여러 곳에서 정보 열매를 하나씩 따는 검색'이다.

NoteBase

5. 메타데이터 입력 및 전송 표준(METS: Metadata Encoding and Transmission Standard)의 7가지 섹션으로 볼 때 '에이전트명, 문서의 생성 및 갱신 시간, 문서상태' 등의 정보를 포함하고 있는 부분은?

① METS 헤더
② 기술 메타데이터 섹션
③ 관리 메타데이터 섹션
④ 구조맵 섹션
⑤ 행위 섹션

정답 1번

출제 포인트

METS를 구성하는 주요 섹션 중 에이전트명, METS 문서의 생성·갱신 시간, 문서 상태를 기록하는 섹션을 구별하는 문제이다.

핵심 개념

METS 헤더(metsHdr)는 METS 문서 자체에 관한 관리 정보를 담는 머리말 영역이다. 대표적으로 에이전트 정보, 생성일시(CREATEDATE), 최종 수정일시(LASTMODDATE), 레코드 상태(RECORDSTATUS)를 기록한다.
기술 메타데이터는 디지털 객체의 기술적 특성을, 관리 메타데이터 섹션은 기술·권리·출처·디지털 보존 이력 등의 관리 정보를 묶어 담는다. 구조맵은 객체의 논리적·물리적 구조를 나타내고, 행위 섹션은 콘텐츠에 적용할 동작과 실행 수단을 연결한다.

선지별 해설

① O — METS 헤더는 에이전트명과 METS 문서의 생성일시·최종 수정일시·레코드 상태를 포함하므로 제시된 설명에 해당한다.
② X — 틀린 표현은 '기술 메타데이터 섹션'이다. 기술 메타데이터는 파일의 형식 등 디지털 객체의 기술적 특성을 기술하며, 제시된 정보는 METS 헤더에 기록한다.
③ X — 틀린 표현은 '관리 메타데이터 섹션'이다. 관리 메타데이터 섹션은 기술 메타데이터, 지적재산권 메타데이터, 출처 메타데이터, 디지털 출처·보존 이력 메타데이터 등을 묶는 영역이며, 제시된 정보의 올바른 섹션은 METS 헤더이다.
④ X — 틀린 표현은 '구조맵 섹션'이다. 구조맵 섹션은 디지털 객체의 논리적·물리적 계층 구조와 각 구성 부분에 대응하는 파일을 나타내며, 제시된 정보는 METS 헤더에 기록한다.
⑤ X — 틀린 표현은 '행위 섹션'이다. 행위 섹션은 디지털 객체의 콘텐츠에 적용되는 동작과 그 실행 수단을 연결하며, 제시된 정보의 올바른 섹션은 METS 헤더이다.

NoteBase

오답 포인트

‘생성 및 갱신 시간’이라는 말만 보고 파일의 기술 정보나 관리 메타데이터를 고르기 쉽다. 그러나 문제의 시간과 상태는 개별 콘텐츠 파일이 아니라 METS 문서 자체의 생성·수정·상태 정보이며, 에이전트 정보와 함께 METS 헤더에 놓인다.

다시 나오면 이렇게 판단

에이전트, METS 문서의 생성일시·최종 수정일시, 레코드 상태가 함께 제시되면 METS 헤더로 판단한다. 객체의 계층은 구조맵, 파일의 기술적 특성은 기술 메타데이터, 실행 동작의 연결은 행위 섹션으로 구별한다.

암기 포인트

METS 문서의 사람·시간·상태는 헤더, 객체의 계층은 구조맵, 실행 기능은 행위 섹션이다.

NoteBase

6. <보기>는 DOI(Digital Object Identifier)의 사례이다. 이에 대한 설명으로 옳지 않은 것은?

<보 기>

http://doi.org/10.12345/kcidoi.2017.1.1.53

① 디지털 객체를 유일하게 식별하고 영구적인 접근을 보장한다.
② "10.12345"는 접두부(Prefix)라 한다.
③ "doi.org"는 디렉토리 관리자이다.
④ "10.12345"에서 "12345"는 등록기관번호이다.
⑤ "kcidoi.2017.1.1.53"은 접미부(Suffix)로 "10"기관이 부여한다.

정답 5번

출제 포인트

DOI의 기능과 DOI 문자열을 구성하는 접두부(Prefix), 접미부(Suffix), 그리고 doi.org의 역할을 구별하는 문제이다.

핵심 개념

DOI는 디지털 객체에 부여하는 영구적이고 고유한 식별자이다. 예시의 DOI 본체는 10.12345/kcidoi.2017.1.1.53이며, 슬래시 앞의 10.12345는 접두부, 슬래시 뒤의 kcidoi.2017.1.1.53은 접미부이다. 접두부의 10은 DOI임을 나타내는 디렉터리 지시자이고, 12345는 등록기관을 식별하는 번호이다. 접미부는 해당 접두부를 사용하는 등록자가 객체를 구별하도록 정한다. doi.org는 DOI를 해당 객체의 현재 인터넷 주소로 연결하는 해석 서비스의 도메인이다.

선지별 해설

① O — DOI는 디지털 객체를 고유하게 식별하며, 객체의 인터넷 주소가 바뀌더라도 DOI에 연결된 정보를 갱신함으로써 지속적인 접근을 지원한다.
② O — 10.12345는 슬래시 앞에 위치한 접두부(Prefix)이다.
③ O — doi.org는 DOI를 실제 객체의 현재 위치로 연결해 주는 DOI 해석 서비스의 도메인이므로, 제시된 표현에서 디렉토리 관리자 역할에 해당한다.
④ O — 접두부 10.12345에서 10은 DOI 디렉터리 지시자이고, 12345는 등록기관을 식별하는 번호이다.
⑤ X — 틀린 표현은 '접미부(Suffix)로 10 기관이 부여한다'이다. kcidoi.2017.1.1.53은 접미부인 것은 맞지만, 접미부는 10이라는 기관이 부여하는 것이 아니라 해당 DOI 등록자가 객체를 고유하게 구별하도록 정한다. 올바르게는 'kcidoi.2017.1.1.53은 접미부이며, 등록자가 객체 식별을 위해 부여한다'이다.

NoteBase

오답 포인트

⑤는 접미부의 범위는 맞게 제시하고 부여 주체를 바꾸어 놓았다. 숫자 10은 기관명이 아니라 DOI 체계임을 나타내는 디렉터리 지시자라는 점이 함정이다.

다시 나오면 이렇게 판단

DOI 본체에서 슬래시를 기준으로 앞은 접두부, 뒤는 접미부로 나눈다. 접두부는 DOI 체계와 등록기관을 식별하고, 접미부는 등록자가 개별 객체를 식별하도록 정하며, doi.org는 DOI를 현재 위치로 연결한다.

암기 포인트

슬래시 앞은 Prefix, 뒤는 Suffix이다. 10은 기관이 아니라 DOI 지시자이고, 접미부의 부여 주체는 등록자이다.

NoteBase

7. <보기> 중 이용자가 가져온 특정 논문을 활용하여 탐색의 범위를 넓히기 위해 취할 수 있는 효과적인 방법을 모두 고른 것은?

<보 기>

ㄱ. 그 논문의 참고문헌을 Chaining한다.

ㄴ. 그 논문의 해당 주제 데이터베이스를 Citation Pearl Growing 기법을 활용한 탐색을 한다.

ㄷ. A&HCI, SSCI, SCI와 같은 인용색인데이터베이스에서 그 논문을 찾아서 인용하거나 인용된 논문들을 탐색한다.

ㄹ. 그 논문에 사용된 주제어를 사용하여 Successive Fraction 기법을 활용한 탐색을 한다.

① ㄱ, ㄴ ② ㄴ, ㄷ ③ ㄷ, ㄹ ④ ㄱ, ㄴ, ㄷ ⑤ ㄱ, ㄴ, ㄷ, ㄹ

정답 4번

출제 포인트

이미 알고 있는 특정 논문을 출발점으로 삼아 관련 문헌을 더 넓게 찾는 탐색 기법과, 검색 결과를 단계적으로 좁히는 기법을 구별하는 문제이다.

핵심 개념

Chaining은 한 논문의 참고문헌이나 인용 관계를 따라 관련 문헌을 연쇄적으로 찾는 방법이다.
Citation Pearl Growing은 적합한 핵심 논문을 진주처럼 출발점으로 삼아 그 논문의 주제어·색인어·인용정보 등을 활용해 관련 문헌을 확장하는 방법이다.
인용색인 탐색은 특정 논문이 인용한 문헌과 그 논문을 인용한 후속 문헌을 모두 추적할 수 있으므로 탐색 범위를 넓히는 데 효과적이다.
Successive Fraction은 넓은 검색 결과에 조건이나 개념을 차례로 추가하여 결과 집합을 단계적으로 줄이는 방법이다. 따라서 범위 확대가 아니라 범위 축소에 해당한다.

NoteBase

선지별 해설

① X — ㄱ과 ㄴ은 모두 범위를 넓히는 방법이지만, 유효한 ㄷ이 빠져 있다. 올바르게는 ㄱ, ㄴ, ㄷ을 모두 포함해야 한다.
② X — ㄴ과 ㄷ은 맞지만, 참고문헌을 연쇄적으로 추적하는 ㄱ이 빠져 있다. 올바르게는 ㄱ, ㄴ, ㄷ을 모두 포함해야 한다.
③ X — ㄷ은 맞지만 ㄹ은 검색 범위를 단계적으로 좁히는 Successive Fraction을 잘못 포함한 것이다. 또한 유효한 ㄱ과 ㄴ이 빠져 있다. 올바르게는 ㄱ, ㄴ, ㄷ이다.
④ O — ㄱ의 Chaining, ㄴ의 Citation Pearl Growing, ㄷ의 인용색인 탐색은 모두 특정 논문을 단서로 관련 문헌을 확장하는 효과적인 방법이다.
⑤ X — ㄱ, ㄴ, ㄷ은 맞지만 ㄹ을 포함한 부분이 틀리다. ㄹ의 '주제어를 사용한 Successive Fraction 탐색'은 조건을 연속적으로 적용해 검색 집합을 좁히는 방식이므로 제외해야 하며, 올바르게는 ㄱ, ㄴ, ㄷ이다.

오답 포인트

ㄹ에 '그 논문에 사용된 주제어'라는 표현이 있어 관련 문헌을 확장하는 기법처럼 보이는 것이 함정이다. 그러나 기법의 성격은 사용한 단서보다 검색 집합의 변화 방향으로 판단해야 한다. Chaining·Citation Pearl Growing·인용색인 탐색은 관련 문헌을 확장하고, Successive Fraction은 검색 결과를 단계적으로 축소한다.

다시 나오면 이렇게 판단

특정 적합 문헌의 참고문헌·인용문헌·색인어를 따라가면 탐색 확장으로 판단한다. 반대로 검색 조건을 차례로 더해 결과 집합을 잘게 나누거나 줄이면 Successive Fraction에 의한 탐색 축소로 판단한다.

암기 포인트

Chaining·Pearl Growing·인용색인은 넓히고, Successive Fraction은 좁힌다.

NoteBase

8. 시맨틱 웹과 온톨로지에 대한 설명으로 옳은 것은?

① 시맨틱 웹의 계층구조는 URI-XML-Logic-RDF-OWL-Proof-Trust 계층의 순서대로 이루어진다.

② N. Guarino는 온톨로지의 유형을 일반성 수준에 따라 top-level ontology, domain ontology, task ontology, application ontology로 구분하였다.

③ 온톨로지 언어의 구성요소는 클래스(class), 인스턴스(instance), 관계(relation), 속성(property), 어커런스(occurrence)이다.

④ 토픽맵은 W3C가 제안한 온톨로지 언어로서 처음 제안된 토픽맵 표준 스펙은 HTML구조의 HyTM 언어이다.

⑤ 온톨로지를 표현하는 언어인 RDF(S)는 유럽 On-To-Knowledge 프로젝트에 의해 개발되었다.

정답 2번

출제 포인트

시맨틱 웹의 계층 구조와 온톨로지의 유형·구성 요소, 그리고 Topic Maps와 RDF(S)의 표준화 주체를 구별하는 문제이다.

핵심 개념

시맨틱 웹: 웹 자원에 기계가 해석할 수 있는 의미를 부여하고, URI·XML·RDF/RDFS·OWL·논리·증명·신뢰 등의 기술을 계층적으로 결합하는 웹이다.
온톨로지: 특정 영역의 개념과 개념 사이의 관계를 명시적으로 표현한 지식 모델이다.
Guarino의 분류: 일반성 수준에 따라 최상위 온톨로지, 도메인 온톨로지, 과업 온톨로지, 응용 온톨로지로 구분한다.

NoteBase

선지별 해설

① X — ‘Logic-RDF-OWL’의 순서가 틀렸다. 시맨틱 웹 계층을 단순화하면 아래에서 위로 URI → XML → RDF/RDFS → OWL → Logic → Proof → Trust의 순서로 이해한다. 따라서 Logic은 RDF와 OWL보다 위에 놓인다.
② O — Guarino는 온톨로지를 일반성 수준에 따라 top-level ontology, domain ontology, task ontology, application ontology로 구분하였다. 제시된 네 유형과 분류 기준이 모두 맞다.
③ X — ‘속성(property), 어커런스(occurrence)’를 온톨로지 언어의 대표적 5대 구성 요소로 나열한 부분이 부정확하다. 대표적 구성 요소는 클래스, 인스턴스, 관계, 함수, 공리이며, occurrence는 Topic Maps에서 토픽과 관련 정보자원을 연결할 때 사용하는 용어이다.
④ X — ‘W3C가 제안’과 ‘HTML 구조의 HyTM’이 틀렸다. Topic Maps는 ISO 계열의 표준이며, 초기 표준 표현인 HyTM은 HTML이 아니라 HyTime을 바탕으로 한 SGML 아키텍처이다.
⑤ X — ‘RDF(S)는 On-To-Knowledge 프로젝트에 의해 개발되었다’가 틀렸다. RDF와 RDFS는 W3C에서 표준화하였으며, 유럽 On-To-Knowledge 프로젝트와 관련된 온톨로지 언어는 OIL이다.

오답 포인트

③의 occurrence는 Topic Maps의 용어인데 온톨로지 구성 요소처럼 섞여 있다. 또한 ④와 ⑤는 표준화 기구와 프로젝트를 서로 바꾸어 놓았고, ①은 시맨틱 웹 계층에서 Logic을 RDF·OWL보다 아래에 둔 것이 함정이다.

다시 나오면 이렇게 판단

계층 순서 문제에서는 RDF/RDFS와 OWL이 의미 표현을 담당하고 Logic이 그 위에서 추론을 담당한다는 관계를 확인한다. 출처 문제에서는 Topic Maps는 ISO, RDF(S)는 W3C, OIL은 On-To-Knowledge로 연결하고, 온톨로지 유형 문제에서는 Guarino의 ‘최상위-도메인-과업-응용’ 분류를 기준으로 판단한다.

암기 포인트

Guarino는 최상위-도메인-과업-응용, Topic Maps는 ISO, RDF(S)는 W3C, OIL은 On-To-Knowledge이다.

NoteBase

9. <보기>는 가상의 RDF/XML 기반의 메타데이터 일부분이다. 이에 대한 설명으로 옳은 것은?

<보 기>

```
<?xml version="1.0"?>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description㉠rdf:about="http://www.ilrt.bristol.ac.uk/people/cmdjb/">
  <dc:title> Guidance on expressing the Dublin core within the RDF </dc:title>
㉡<dc:creator> Eric Miller </dc:creator>
  <dc:subject> Dublin Core; Resource Description Framework; RDF</dc:subject>
  <dc:publisher> Dublin Core Matadata Initiative </dc:publisher>
  <dc:data> 1999-07-10 </dc:data>
  <dc:format> text/html </dc:format>
 </rdf:Description>
</rdf:RDF>
```

① ㉠은 기술의 대상이 되는 정보자원이 URI로 식별되는 경우에 사용한다.

② 이 메타데이터에는 네임스페이스를 사용하지 않았다.

③ Dublin Core 세분화 요소를 이용한 한정어를 사용하였다.

④ ㉡은 지적소유권 기술요소로서 MODS의 상위요소 'author'에 해당한다.

⑤ XML 구문을 사용한다는 XML 선언이 없다.

정답 1번

출제 포인트

RDF/XML에서 자원을 URI로 식별하는 속성, XML 네임스페이스, 단순 Dublin Core 요소와 MODS 요소의 대응을 구별하는 문제이다.

핵심 개념

RDF/XML은 자원에 관한 문장을 XML 문법으로 표현하는 방식이다. rdf:about은 rdf:Description이 설명하는 자원을 URI로 지정한다. xmlns:rdf와 xmlns:dc는 각각 rdf와 dc 접두사가 가리키는 네임스페이스를 선언한다. 보기의 dc:title, dc:creator, dc:subject 등은 세분화 요소가 붙지 않은 단순 Dublin Core 요소이며, MODS에서는 저자 정보를 상위요소 name과 그 하위 역할 정보로 표현한다.

NoteBase

선지별 해설

① O — ㉠의 rdf:about은 rdf:Description의 기술 대상이 되는 정보자원을 URI로 식별하는 속성이다. 따라서 선지의 설명이 옳다.
② X — 틀린 표현은 '네임스페이스를 사용하지 않았다'이다. 실제로는 xmlns:rdf와 xmlns:dc로 RDF 및 Dublin Core 네임스페이스를 선언하고 rdf:, dc: 접두사를 사용하였다.
③ X — 틀린 표현은 'Dublin Core 세분화 요소를 이용한 한정어를 사용하였다'이다. 실제로는 dc:title, dc:creator, dc:subject 등 단순 Dublin Core 요소만 사용하였고, 세분화 요소나 인코딩 스킴 한정어는 나타나지 않는다.
④ X — 틀린 표현은 'MODS의 상위요소 author에 해당한다'이다. ㉡ dc:creator는 자원의 작성자·기여 책임 주체를 나타내지만, MODS에는 author라는 상위요소가 없다. 실제로 MODS의 저자 정보는 상위요소 name 아래에서 role 또는 roleTerm으로 author 역할을 표시한다.
⑤ X — 틀린 표현은 'XML 선언이 없다'이다. 실제로 보기 첫 줄에 <?xml version="1.0"?>이라는 XML 선언이 있다.

오답 포인트

③은 dc:subject 값에 여러 주제어가 들어 있다는 사실을 요소 세분화와 혼동하게 만든다. 값에 주제어가 여러 개 있는 것과 한정어로 요소를 세분화하는 것은 다르다. ④는 Dublin Core의 creator를 MODS의 author라는 상위요소와 곧바로 대응시키는 표현이 함정이다.

다시 나오면 이렇게 판단

먼저 xmlns 선언으로 네임스페이스 사용 여부를 확인하고, rdf:about 뒤의 URI가 기술 대상 자원을 가리키는지 본다. 이어 요소 이름에 세분화 표현이 실제로 있는지 확인하며, 서로 다른 메타데이터 스키마의 요소명을 동일하다고 단정하지 않는다.

암기 포인트

rdf:about은 '무엇을 설명하는가'를 URI로 지정하고, xmlns는 접두사의 소속 네임스페이스를 선언한다. MODS의 저자는 author 상위요소가 아니라 name과 역할 정보로 표현한다.

NoteBase

10. <보기>는 "정보검색시스템의 역사"라는 문장을 입력하여 ㉠, ㉡과 같은 색인어를 추출한 자동색인기법에 대한 사례이다. 이에 대한 설명으로 옳은 것은?

<보 기>

입력문장	추출된 색인어
정보검색시스템의 역사	㉠ 정보검색시스템, 역사
	㉡ 정보, 검색, 시스템, 역사

① ㉠은 형태소 단위, ㉡은 어절 단위로 색인어를 추출한 것이다.
② ㉠과 같은 색인어 추출 방식의 경우 불용어 제거 작업이 쉽다.
③ ㉡의 색인어 추출 방식은 ㉠의 색인어 추출방식이 갖는 띄어쓰기문제를 해결할 수 있다.
④ ㉡의 색인어 추출 방식은 복합어를 단일어로 분리할 수 없다.
⑤ n-gram은 ㉠과 ㉡의 색인방식을 조합한 것이다.

정답 3번

출제 포인트

어절 단위 색인과 형태소 단위 색인의 추출 결과 및 장단점을 구별하는 문제이다.

핵심 개념

어절 단위 색인은 띄어쓰기를 기준으로 문장을 나누어 색인어를 추출하는 방식이다. 예시의 ㉠처럼 '정보검색시스템'을 하나의 단위로 처리하므로 추출은 단순하지만, 띄어쓰기 오류와 조사·어미 처리에 취약하고 복합어의 구성 단어를 세분하기 어렵다.
형태소 단위 색인은 문장을 의미를 지닌 최소 단위로 분석하여 색인어를 추출하는 방식이다. 예시의 ㉡처럼 '정보검색시스템'을 '정보·검색·시스템'으로 분리할 수 있어 복합어 처리와 색인어 정규화에 유리하다.
n-gram 방식은 문자열을 일정한 길이의 연속된 문자 또는 단어 조각으로 나누는 방식이며, 어절 방식과 형태소 방식의 단순한 결합이 아니다.

NoteBase

선지별 해설

① X — ㉠과 ㉡의 단위를 서로 바꾼 표현이다. 올바르게는 ㉠이 어절 단위, ㉡이 형태소 단위 색인이다.
② X — ㉠과 같은 어절 단위 방식은 조사나 어미가 결합된 형태를 그대로 처리하기 쉬워 불용어를 정확히 분리·제거하는 데 한계가 있다. 올바르게는 형태소 분석을 거쳐 조사·어미 등을 구분하는 방식이 불용어 제거에 더 유리하다.
③ O — ㉡의 형태소 단위 방식은 문장을 형태소로 분석하므로 띄어쓰기 단위에 의존하는 ㉠의 문제를 보완할 수 있다.
④ X — '복합어를 단일어로 분리할 수 없다'는 부분이 틀렸다. 올바르게는 형태소 분석을 통해 '정보검색시스템'을 '정보·검색·시스템'처럼 구성 단어로 분리할 수 있다.
⑤ X — '㉠과 ㉡의 색인 방식을 조합한 것'이라는 설명이 틀렸다. 올바르게는 n-gram이 문자열을 일정한 길이의 연속된 문자 또는 단어 조각으로 분할하여 색인하는 독립적인 방식이다.

오답 포인트

㉠과 ㉡의 명칭을 반대로 연결하거나, 어절 단위 색인의 '추출이 단순하다'는 특징을 '불용어 제거가 쉽다'는 뜻으로 확대하면 틀리기 쉽다. 또한 n-gram은 두 방식의 혼합이 아니라 고정 길이의 연속 조각을 이용하는 별도 방식이다.

다시 나오면 이렇게 판단

띄어쓰기대로 큰 덩어리가 남으면 어절 단위, 복합어가 의미 단위로 나뉘면 형태소 단위로 판단한다. 일정 길이의 연속 조각으로 자르면 n-gram 방식으로 판단한다.

암기 포인트

어절은 띄어쓰기, 형태소는 의미의 최소 단위, n-gram은 일정 길이의 연속 조각이다.

NoteBase

11. 통제어색인의 특징으로 옳지 않은 것은?

① 특정한 개념은 항상 같은 용어로 색인되므로 재현율을 향상시킬 수 있다.

② 용어 간의 관계를 이용하여 일차 선택한 탐색어 외에 다른 탐색어를 추가하여 탐색영역을 확장할 수 있다.

③ 통제어휘는 검색시간을 절약할 수 있으나 색인 생성에 드는 시간과 비용이 자연어색인에 비하여 크다.

④ 용어의 특정성이 높아 주제의 구체적인 표현이 가능하다.

⑤ 전문주제분야의 경우 어휘변화를 반영하기 위해 지속적으로 갱신하여야 한다.

정답 4번

출제 포인트

통제어색인의 장점과 한계를 자연어색인과 비교하여 구별하는 문제이다.

핵심 개념

통제어색인은 미리 정한 표준 용어로 문헌의 주제를 표현하는 색인 방식이다. 동의어와 유사어를 하나의 용어로 통일하고 용어 사이의 관계를 구조화하므로 일관성과 재현율을 높일 수 있다. 다만 새로운 개념이나 매우 구체적인 표현을 즉시 반영하기 어렵고, 통제어휘를 구축하고 계속 갱신하는 데 시간과 비용이 든다.
통제어색인은 용어 통일과 관계 구조화가 강점이고, 자연어색인은 최신성·구체성·표현의 자유가 강점이다.

NoteBase

선지별 해설

① O
특정 개념을 하나의 표준 용어로 통일하여 색인하므로 동의어나 표현 차이로 인한 검색 누락이 줄어 재현율을 높일 수 있다.

② O
상위어·하위어·관련어 같은 용어 간 관계를 활용하면 최초 탐색어에서 다른 적절한 탐색어로 확장하여 탐색영역을 넓힐 수 있다.

③ O
표준 용어와 용어 관계를 이용하면 검색 시 적절한 탐색어를 찾는 시간이 줄어든다. 반면 통제어휘의 구축·유지와 색인어 선정에는 자연어색인보다 많은 시간과 비용이 들 수 있다.

④ X
틀린 표현은 '용어의 특정성이 높아 주제의 구체적인 표현이 가능하다'이다. 통제어색인은 미리 정해진 용어만 사용하므로 새로운 주제나 세부적인 개념을 자유롭게 표현하는 특정성이 자연어색인보다 낮다. 올바르게는 '용어의 특정성이 상대적으로 낮아 주제의 매우 구체적이거나 새로운 표현에 제약이 있다'이다.

⑤ O
전문 분야의 신조어와 개념 변화가 통제어휘에 자동으로 반영되는 것은 아니므로 지속적인 추가·수정·갱신이 필요하다.

오답 포인트

④의 '특정성이 높다'가 함정이다. 통제어색인은 용어를 표준화하므로 검색의 일관성과 재현율은 높지만, 사용 가능한 표현이 제한되어 자연어색인보다 최신 개념이나 세부 주제를 구체적으로 표현하기 어렵다.

다시 나오면 이렇게 판단

통제어색인의 특징은 '표준화·동의어 통제·용어 관계·재현율 향상'으로 판단한다. '최신성·자유로운 표현·높은 특정성'을 장점으로 제시하면 자연어색인의 특징이 섞였는지 확인한다.

암기 포인트

통제어는 통일·관계·재현율에 강하고, 자연어는 최신·구체·자유 표현에 강하다.

NoteBase

12. 정보검색 분야의 발전에 기여한 인물과 업적의 연결이 옳지 않은 것은?

① M. Taube - 유니텀 개념을 사용한 조합색인 기법 창안
② K. Spärck Jones - 인용색인 기반의 정보검색시스템 개발
③ C. N. Mooers - Zatocoding 시스템 개발
④ G. Salton - SMART를 이용한 정보검색 실험
⑤ H. P. Luhn - 컴퓨터 기반의 자동색인 기법 KWIC 제안

정답 2번

출제 포인트

정보검색 발전사에서 주요 인물과 대표 업적이 올바르게 연결되었는지를 판단하는 문제이다.

핵심 정리

M. Taube: 유니텀과 조합색인 — 문헌의 내용을 단일 개념어인 유니텀으로 나타내고 여러 유니텀을 조합하여 검색하는 방식을 발전시켰다.
K. Spärck Jones: 용어 특이성과 역문헌빈도(IDF) — 문헌 집합에서 드물게 나타나는 용어일수록 검색어로서 구별 능력이 크다는 가중 원리를 제시했다.
C. N. Mooers: Zatocoding — 문헌의 주제 특성을 코드로 중첩하여 검색하는 정보검색 시스템을 개발했다.
G. Salton: SMART — 벡터공간모형과 자동 정보검색 연구를 SMART 시스템을 통해 발전시켰다.
H. P. Luhn: KWIC — 표제나 문장에 나타난 핵심어를 중심에 두고 주변 문맥을 함께 제시하는 자동색인 방식을 제안했다.

선지별 해설

① O — M. Taube는 단일 개념어인 유니텀을 조합하여 검색하는 조합색인 기법과 연결된다.
② X — 틀린 부분은 'K. Spärck Jones가 인용색인 기반의 정보검색시스템을 개발했다'는 연결이다. Spärck Jones의 대표 업적은 용어의 특이성을 계량화한 역문헌빈도(IDF) 개념이며, 인용색인의 대표 인물은 Eugene Garfield이다. 올바르게는 'K. Spärck Jones — 용어 특이성 및 역문헌빈도(IDF) 연구'로 고쳐야 한다.
③ O — C. N. Mooers는 주제 개념을 코드화하고 중첩하여 검색하는 Zatocoding 시스템을 개발했다.
④ O — G. Salton은 SMART 시스템을 이용하여 자동 정보검색을 실험하고 벡터공간모형 연구를 발전시켰다.
⑤ O — H. P. Luhn은 핵심어와 그 앞뒤 문맥을 함께 배열하는 KWIC를 제안했으며, 이는 컴퓨터 기반 자동색인의 대표적 방식이다.

NoteBase

출제자가 노린 함정

②의 '인용색인'도 정보검색 분야의 중요한 개념이어서 그럴듯해 보이지만, Spärck Jones의 핵심 업적은 인용관계가 아니라 문헌빈도를 이용한 용어 가중치인 IDF이다. 인물과 정보검색 용어가 모두 익숙하다는 이유만으로 연결하면 틀리기 쉽다.

다시 나오면 이렇게 판단

인물명을 대표 키워드와 일대일로 연결하여 판단한다. Taube는 유니텀·조합색인, Mooers는 Zatocoding, Salton은 SMART·벡터공간모형, Luhn은 KWIC, Spärck Jones는 IDF로 구분한다.

암기 포인트

Taube=Uniterm, Mooers=Zatocoding, Salton=SMART, Luhn=KWIC, Spärck Jones=IDF이다.

NoteBase

13. <보기>는 특정 원저작물에 부여된 CCL 사례이다. 이에 대한 설명으로 옳지 않은 것은?

<보 기>



① 복사 및 배포할 수 있다.
② 상업적 이용이 가능하다.
③ 동일조건 하에 저작물의 변경이 가능하다.
④ 2차적 저작물의 라이선스를 자유롭게 선택할 수 있다.
⑤ 반드시 저작자 및 출처를 표시해야 한다.

정답 4번

출제 포인트

보기의 CCL 표시는 저작자표시(BY)와 동일조건변경허락(SA)의 결합인 CC BY-SA이다. 이 표시는 저작자와 출처를 밝히고, 변경한 저작물을 배포할 때 동일한 라이선스 조건을 적용해야 한다는 뜻이다.

선지별 해설

① O — CC BY-SA 저작물은 라이선스 조건을 지키면 복사하고 배포할 수 있다.
② O — 이 라이선스에는 비영리(NC) 표시가 없으므로 상업적 이용도 가능하다.
③ O — 저작물을 수정·변형할 수 있으며, 변경한 저작물을 공유할 때에는 동일한 CC BY-SA 조건을 적용해야 한다.
④ X — '라이선스를 자유롭게 선택할 수 있다'는 표현이 틀리다. 올바르게는 2차적 저작물을 배포할 때 원저작물과 동일한 CC BY-SA 또는 이에 호환되는 라이선스를 적용해야 한다.
⑤ O — BY는 저작자표시 조건이므로 저작자와 출처 등 요구되는 귀속 정보를 표시해야 한다.

오답 포인트

SA를 단순히 '변경 허용'으로만 이해하면 ④를 놓치기 쉽다. SA는 변경을 허용하면서도 변경한 저작물의 배포 조건을 같은 라이선스로 이어 가게 하는 동일조건변경허락이다. 따라서 이용자가 2차적 저작물의 라이선스를 임의로 고를 수 있는 것은 아니다.

NoteBase

다시 나오면 이렇게 판단

CCL 아이콘을 조건별로 분해해 판단한다. BY는 저작자표시, NC는 비영리, ND는 변경금지, SA는 변경한 저작물에 대한 동일조건 적용을 뜻한다. 보기에 없는 제한은 임의로 추가하지 않는다.

암기 포인트

BY-SA는 ‘출처를 밝히고, 변경본도 같은 조건으로 공유’이다.

NoteBase

14. <보기>의 데이터 요소를 모두 포괄하는 MODS(Metadata Object Description Schema)의 최상위 요소명은?

<보 기>

place, publisher, copyrightDate

① typeOfResource
② accessCondition
③ genre
④ physicalDescription
⑤ originInfo

정답 5번

출제 포인트

MODS의 하위 데이터 요소를 보고 이를 직접 포괄하는 최상위 요소를 판별하는 문제이다.

핵심 개념

originInfo는 자원의 기원과 발행에 관한 정보를 묶는 MODS의 최상위 요소로, place, publisher, copyrightDate 등을 하위 요소로 포함한다.
typeOfResource는 자원의 일반적 유형을, accessCondition은 이용 및 접근 조건을, genre는 자료의 장르나 형식을, physicalDescription은 물리적 형태와 특성을 기술한다.

선지별 해설

① X — typeOfResource는 텍스트·지도·음악 등 자원의 일반적 유형을 나타내는 요소이다. place, publisher, copyrightDate를 포괄한다는 부분이 틀렸으며, 올바르게는 originInfo가 이들을 포괄한다.
② X — accessCondition은 이용 제한, 접근 조건, 권리 관련 조건을 기술하는 요소이다. place, publisher, copyrightDate를 포괄한다는 부분이 틀렸으며, 올바르게는 originInfo가 이들을 포괄한다.
③ X — genre는 자원의 장르나 형식을 나타내는 요소이다. place, publisher, copyrightDate를 포괄한다는 부분이 틀렸으며, 올바르게는 originInfo가 이들을 포괄한다.
④ X — physicalDescription은 형태, 크기, 매체 등 자원의 물리적 특성을 기술하는 요소이다. place, publisher, copyrightDate를 포괄한다는 부분이 틀렸으며, 올바르게는 originInfo가 이들을 포괄한다.
⑤ O — originInfo는 자원의 기원·발행 정보를 묶는 최상위 요소이므로 place, publisher, copyrightDate를 모두 하위 요소로 포함한다.

NoteBase

오답 포인트

copyrightDate만 보고 accessCondition을 고르거나, publisher를 자료의 형태 정보로 오인하여 physicalDescription을 고르기 쉽다. 그러나 세 요소의 공통축은 권리나 물리적 형태가 아니라 자원의 기원·발행 정보이다.

다시 나오면 이렇게 판단

하위 요소들의 공통 의미를 먼저 묶는다. place·publisher·각종 발행 및 생성 날짜처럼 자원의 출처와 발행 이력을 나타내는 항목이면 originInfo로 판단한다.

암기 포인트

장소·발행자·발행 관련 날짜는 originInfo이다.

NoteBase

15. <보기>에서 설명하고 있는 법칙은?

<보 기>

특정 주제 분야에서 많은 수의 논문을 생산하는 저자의 수는 그 분야의 총 저자 수의 제곱근에 가까우며, 그들의 논문은 그 분야 출판물의 절반을 차지한다는 법칙

① de Solla Price의 법칙
② Pareto의 법칙
③ Zipf의 법칙
④ Willis의 법칙
⑤ Long Tail의 법칙

정답 1번

출제 포인트

이 문제는 논문 생산의 집중 현상을 설명하는 문장과 대표 정보계량 법칙을 연결하는 문제이다.

핵심 정리

de Solla Price의 법칙: 한 분야의 전체 저자 수의 제곱근에 해당하는 소수의 생산적인 저자들이 전체 논문의 약 절반을 생산한다.
Pareto의 법칙: 대체로 소수인 20%의 원인이 결과의 80%를 만들어 낸다는 80 대 20 법칙이다.
Zipf의 법칙: 단어 등의 사용 빈도는 그 빈도 순위에 대체로 반비례한다.
Willis의 법칙: 생물 분류군의 연령과 종의 수·지리적 분포 범위 사이의 관계를 설명하며, 그 분포 법칙은 정보계량 연구에도 응용된다.
Long Tail의 법칙: 개별 수요는 적지만 종류가 매우 많은 비인기 항목들의 누적 수요가 큰 시장을 형성할 수 있다는 개념이다.

NoteBase

선지별 해설

① O — 보기의 '전체 저자 수의 제곱근에 가까운 저자들이 논문의 절반을 생산한다'는 내용은 de Solla Price의 제곱근 법칙에 해당한다.
② X — 틀린 부분은 보기의 현상을 Pareto의 법칙과 연결한 것이다. 올바르게는 Pareto의 법칙은 소수의 원인 약 20%가 결과 약 80%를 낳는다는 80 대 20 법칙이다.
③ X — 틀린 부분은 보기의 현상을 Zipf의 법칙과 연결한 것이다. 올바르게는 Zipf의 법칙은 단어 등의 빈도가 빈도 순위에 대체로 반비례한다는 법칙이다.
④ X — 틀린 부분은 보기의 현상을 Willis의 법칙과 연결한 것이다. 올바르게는 Willis의 법칙은 생물 분류군의 연령과 종의 수·지리적 분포 범위 사이의 관계를 다루며 정보계량 분야에도 응용되는 법칙이다.
⑤ X — 틀린 부분은 보기의 현상을 Long Tail의 법칙과 연결한 것이다. 올바르게는 Long Tail은 개별 수요가 적은 다수의 비인기 항목이 누적으로 상당한 수요를 이룰 수 있다는 개념이다.

출제자가 노린 함정

Pareto·Zipf·Long Tail도 모두 소수에 집중되거나 한쪽 꼬리가 긴 불균등 분포와 관련되어 있어 혼동하기 쉽다. 이 문항의 결정적 표현은 '저자 수의 제곱근'과 '논문의 절반'이며, 이 두 표현이 함께 나오면 de Solla Price의 법칙이다.

다시 나오면 이렇게 판단

법칙의 이름보다 수치와 대상의 조합으로 판단한다. '저자 수의 제곱근·논문 절반'은 Price, '20%·80%'는 Pareto, '빈도·순위의 반비례'는 Zipf, '다수의 틈새 항목의 누적 수요'는 Long Tail로 연결한다.

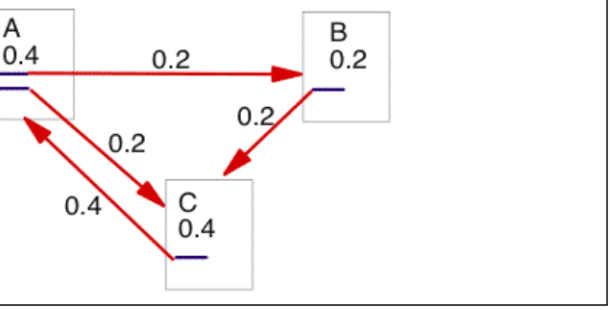
암기 포인트

Price는 제곱근 저자가 절반의 논문을 생산한다.

NoteBase

16. <보기>는 구글 검색엔진의 페이지랭크 알고리즘을 설명하기 위해서 웹페이지 A, B, C 간의 링크구조를 제시한 것이다. 이에 대한 설명으로 옳은 것은?

<보 기>



① 웹페이지 B의 페이지랭크 값이 가장 크다.

② 웹페이지 C는 전진(Forward) 링크가 2건이고, 0.4의 가중치 값을 웹페이지 A로부터 받는다.

③ 웹페이지 A는 후진(Backward) 링크가 1건이고, 웹페이지 C로부터 가중치 0.4의 값을 받는다.

④ 웹페이지 A와 웹페이지 C의 전진(Forward) 링크 수는 동일하다.

⑤ 웹페이지 A와 웹페이지 C의 페이지랭크 값은 차이가 있다.

정답 3번

출제 포인트

이 문제는 링크 화살표의 방향을 기준으로 각 웹페이지의 전진 링크와 후진 링크를 구분하고, 도식에 표시된 페이지랭크 값과 링크 가중치를 정확히 읽는 문제이다. 전진 링크는 해당 페이지에서 다른 페이지로 나가는 링크이고, 후진 링크는 다른 페이지에서 해당 페이지로 들어오는 링크이다.

NoteBase

선지별 해설

① X — 웹페이지 B의 페이지랭크 값은 0.2이고, A와 C는 각각 0.4이므로 B의 값이 가장 크다는 설명은 틀리다. 올바르게는 B의 페이지랭크 값이 가장 작고, A와 C의 값이 공동으로 가장 크다.

② X — C에서 출발하는 화살표는 C→A 한 개뿐이며, 그 링크의 가중치는 0.4이다. A→C와 B→C는 C로 들어오는 후진 링크이다. 올바르게는 C의 전진 링크가 1건이고, C는 가중치 0.4의 값을 A에 전달한다.

③ O — A로 들어오는 화살표는 C→A 한 개이므로 A의 후진 링크는 1건이다. 또한 이 링크에 표시된 가중치가 0.4이므로 A는 C로부터 가중치 0.4의 값을 받는다.

④ X — A에서 나가는 전진 링크는 A→B와 A→C의 2건이지만, C에서 나가는 전진 링크는 C→A의 1건이다. 올바르게는 A의 전진 링크 수가 2건, C의 전진 링크 수가 1건이므로 서로 다르다.

⑤ X — 도식에 표시된 A와 C의 페이지랭크 값은 모두 0.4이다. 올바르게는 A와 C의 페이지랭크 값이 같아 차이가 없다.

오답 포인트

화살표가 페이지에 들어오면 후진 링크, 페이지에서 나가면 전진 링크라는 기준을 뒤집기 쉽다. 특히 C에는 A와 B에서 들어오는 링크가 2건 있지만, 이는 C의 후진 링크이며 C의 전진 링크는 C→A 한 개뿐이다. 또한 상자 안의 수치는 페이지랭크 값이고 화살표 옆의 수치는 링크 가중치이므로 서로 구분해야 한다.

다시 나오면 이렇게 판단

판단할 페이지를 기준으로 화살표 꼬리가 붙어 밖으로 나가면 전진 링크, 화살표 머리가 들어오면 후진 링크로 센다. 그다음 상자 안의 페이지랭크 값과 화살표 옆의 링크 가중치를 각각 따로 확인한다.

암기 포인트

나가는 화살표는 전진 링크, 들어오는 화살표는 후진 링크이다. A는 전진 2건·후진 1건, C는 전진 1건·후진 2건이며 A와 C의 페이지랭크 값은 모두 0.4이다.

NoteBase

17. 한국학술지 인용색인 영향력지수(KCI IF)를 제공하고 있는 기관은?

① 한국연구재단
② 국립중앙도서관
③ 국회도서관
④ 한국과학기술정보연구원
⑤ 한국교육학술정보원

정답 1번

출제 포인트

한국학술지인용색인(KCI)과 KCI 영향력지수(KCI IF)를 제공하는 기관을 구별하는 문제이다.

핵심 개념

KCI는 한국연구재단이 운영하는 국내 학술지 인용정보 서비스이다. 국내 학술지와 논문의 서지·인용 정보를 제공하며, 인용 정보를 바탕으로 학술지의 영향력지수도 산출·제공한다.

선지별 해설

① O — 한국연구재단은 KCI를 운영하며 국내 학술지의 인용정보와 KCI 영향력지수를 제공한다.
② X — 틀린 기관명은 '국립중앙도서관'이다. 국립중앙도서관은 국가 대표 도서관으로서 국내 자료의 수집·보존과 이용 서비스를 담당하며, 올바르게는 KCI IF 제공 기관이 한국연구재단이다.
③ X — 틀린 기관명은 '국회도서관'이다. 국회도서관은 국회의 입법 활동을 위한 정보 지원과 도서관 서비스를 담당하며, 올바르게는 KCI IF 제공 기관이 한국연구재단이다.
④ X — 틀린 기관명은 '한국과학기술정보연구원'이다. 한국과학기술정보연구원은 과학기술 정보와 연구 인프라 서비스를 제공하는 기관이며, 올바르게는 KCI IF 제공 기관이 한국연구재단이다.
⑤ X — 틀린 기관명은 '한국교육학술정보원'이다. 한국교육학술정보원은 교육·학술정보 서비스를 지원하며, 올바르게는 KCI IF 제공 기관이 한국연구재단이다.

오답 포인트

선지의 기관들이 모두 도서관·학술정보와 관련되어 있어 혼동하기 쉽다. 이 문항에서는 일반적인 자료 제공 기관이 아니라 'KCI'라는 고유 서비스의 운영 기관을 묻는다는 점이 핵심이다.

다시 나오면 이렇게 판단

학술정보 서비스의 운영 기관을 묻는 문항에서는 서비스 약칭과 대표 기관을 연결한다. KCI·KCI IF가 제시되면 한국연구재단을 선택한다.

NoteBase

암기 포인트

KCI와 KCI IF는 한국연구재단이다.

NoteBase

18. ISO 9126 표준에서는 이용자 인터페이스의 사용성 평가를 위하여 다섯 가지의 요소를 정의하고 있는데, 이 요소에 해당되지 않는 것은?

① supportability
② operability
③ attractiveness
④ understandability
⑤ compliance

정답 1번

출제 포인트

ISO 9126에서 사용성(usability)을 구성하는 다섯 가지 하위 요소를 구별하는 문제이다.

핵심 개념

사용성은 사용자가 소프트웨어를 이해하고, 배우고, 조작하며, 매력적으로 느낄 수 있는 정도와 사용성 관련 표준·규정의 준수 정도를 나타내는 품질 특성이다.
다섯 가지 하위 요소는 이해성(understandability), 학습성(learnability), 운용성(operability), 매력성(attractiveness), 사용성 준수성(usability compliance)이다.

선지별 해설

① X — supportability는 ISO 9126의 사용성 하위 요소가 아니다. 틀린 표현은 supportability이며, 올바르게는 다섯 요소 중 빠진 learnability(학습성)가 들어가야 한다.
② O — operability(운용성)는 사용자가 소프트웨어를 조작하고 제어할 수 있는 정도이므로 사용성의 하위 요소이다.
③ O — attractiveness(매력성)는 소프트웨어가 사용자에게 호감과 매력을 주는 정도이므로 사용성의 하위 요소이다.
④ O — understandability(이해성)는 사용자가 소프트웨어의 용도와 사용 방법을 이해할 수 있는 정도이므로 사용성의 하위 요소이다.
⑤ O — compliance는 여기서 usability compliance(사용성 준수성)를 뜻하며, 사용성에 관한 표준·관례·규정을 준수하는 정도이므로 사용성의 하위 요소이다.

오답 포인트

①의 supportability가 소프트웨어 품질과 관련된 자연스러운 용어처럼 보이는 점이 함정이다. 그러나 ISO 9126의 사용성 다섯 요소에는 supportability가 아니라 learnability가 포함된다. 또한 ⑤의 compliance는 문맥상 독립된 품질 특성이 아니라 usability compliance를 가리킨다.

NoteBase

다시 나오면 이렇게 판단

사용성 하위 요소를 이해·학습·운용·매력·준수의 다섯 축으로 대조한다. 선지 중 이 다섯 축에 대응하지 않는 용어를 고르면 된다.

암기 포인트

사용성은 이해성·학습성·운용성·매력성·준수성이다. '이·학·운·매·준'으로 기억한다.

NoteBase

19. 정보검색평가에 대한 연구자와 발표 이론의 연결이 옳지 않은 것은?

① S. P. Harter - 심리적 적합성 이론(Psychological Theory of Relevance)
② W. S. Cooper - 논리적 적합성과 유용성 이론(Logical Relevance & Utility Theory)
③ T. D. Wilson - 상황 적합성 이론(Situational Relevance Theory)
④ C. W. Cleverdon - 통합 적합성 이론(Aggregate Relevance Theory)
⑤ F. W. Lancaster - 적합성/필요성 관계 이론(Relevance/Pertinence Relation Theory)

정답 4번

출제 포인트

정보검색에서 적합성을 설명한 연구자와 대표 이론을 정확히 연결하는 문제이다.

핵심 정리

S. P. Harter: 심리적 적합성 — 검색된 정보가 이용자의 심리·인지 상태에 변화를 일으키는지를 중시한다.
W. S. Cooper: 논리적 적합성과 유용성 — 적합성을 논리적 함의 관계로 설명하고, 실제 이용자에게 주는 궁극적 유용성과 구별한다.
T. D. Wilson: 상황 적합성 — 정보가 특정 이용자의 문제 상황과 세계관에 어떤 의미와 변화를 주는지를 중시한다.
C. W. Cleverdon: Cranfield 실험 — 통제된 테스트 컬렉션에서 검색시스템의 성능을 평가한 시스템 중심 연구와 연결된다.
F. W. Lancaster: 적합성·필요성 관계 — 적합성은 문헌과 표현된 요구의 관계, 필요성은 문헌과 실제 정보요구의 관계로 구분한다.

선지별 해설

① O — Harter는 정보가 이용자의 인지 상태를 변화시키는가에 초점을 둔 심리적 적합성을 제시하였다.
② O — Cooper는 정보와 요구 사이의 논리적 관련성을 설명하는 논리적 적합성과, 이용자에게 실제로 얼마나 유용한지를 나타내는 유용성을 논의하였다.
③ O — Wilson은 정보가 특정 개인의 문제 상황과 인식에 미치는 관계를 상황 적합성으로 설명하였다.
④ X — 틀린 부분은 'Cleverdon—통합 적합성 이론'이라는 연결이다. 올바르게는 Cleverdon을 Cranfield 실험과 검색시스템의 성능을 측정하는 시스템 중심 평가에 연결해야 한다.
⑤ O — Lancaster는 문헌과 표현된 요구 사이의 적합성, 문헌과 이용자의 실제 정보요구 사이의 필요성을 구분하여 그 관계를 설명하였다.

NoteBase

출제자가 노린 함정

Cleverdon이 정보검색 평가의 대표 연구자라는 사실 때문에 '통합 적합성 이론'도 그가 발표한 이론처럼 보이게 한 것이 함정이다. Cleverdon의 대표 연결어는 통합 적합성이 아니라 Cranfield 실험과 시스템 성능 평가이다.

다시 나오면 이렇게 판단

학자 이름을 대표 연결어로 먼저 대응한다. Harter는 심리, Cooper는 논리·유용성, Wilson은 상황, Cleverdon은 Cranfield 실험, Lancaster는 적합성·필요성 구분으로 판단한다.

암기 포인트

하터-심리, 쿠퍼-논리·유용성, 윌슨-상황, 클레버던-크랜필드, 랭커스터-적합성·필요성이다.

NoteBase

20. DC(Dublin Core)의 요소와 그에 대응하는 인코딩 스킴(encoding scheme)의 연결이 옳지 않은 것은?

① Subject - DDC
② Date - W3CDTF
③ Source - URI
④ Identifier - URI
⑤ Relation - DCMI Point

정답 5번

출제 포인트

DC(Dublin Core) 요소에 알맞은 인코딩 스킴이 연결되었는지를 묻는 문제이다.

핵심 개념

인코딩 스킴은 메타데이터 요소의 값을 일정한 분류표·형식·식별 체계에 따라 표현하도록 하는 기준이다.
DDC는 주제 분류, W3CDTF는 날짜와 시간, URI는 자원의 식별, DCMI Point는 지리 좌표로 나타내는 공간 범위에 사용된다.
따라서 DCMI Point의 연결 대상은 Relation이 아니라 Coverage이다.

선지별 해설

① O — Subject는 자원의 주제를 나타내며, DDC는 주제를 듀이십진분류법의 분류기호로 표현하는 인코딩 스킴이므로 연결이 옳다.
② O — Date는 자원과 관련된 날짜를 나타내며, W3CDTF는 ISO 8601에 기초하여 날짜와 시간을 표현하는 스킴이므로 연결이 옳다.
③ O — Source는 현재 자원이 유래한 원천 자원을 나타내며, 그 자원을 URI로 식별할 수 있으므로 연결이 옳다.
④ O — Identifier는 자원을 명확하게 식별하는 요소이며, URI는 자원의 식별값을 표현하는 스킴이므로 연결이 옳다.
⑤ X — 틀린 부분은 'Relation - DCMI Point'이다. DCMI Point는 지리 좌표로 공간상의 한 지점을 표현하므로 Relation이 아니라 Coverage에 연결된다. 올바른 연결은 'Coverage - DCMI Point'이다. Relation의 관련 자원은 URI로 식별할 수 있다.

오답 포인트

⑤의 Point를 두 자원 사이의 '관계 지점'처럼 해석하면 Relation과 혼동하기 쉽다. DCMI Point의 Point는 관계가 아니라 위도·경도 등 지리 좌표로 표시되는 공간상의 지점을 뜻한다.

NoteBase

다시 나오면 이렇게 판단

스킴의 이름이 무엇을 표현하는지 먼저 대응시킨다. 분류기호는 Subject, 날짜 형식은 Date, 자원 식별자는 Source·Identifier·Relation, 지리 좌표는 Coverage에 대응한다.

암기 포인트

DDC-Subject, W3CDTF-Date, URI-식별, DCMI Point-Coverage로 기억한다.

NoteBase