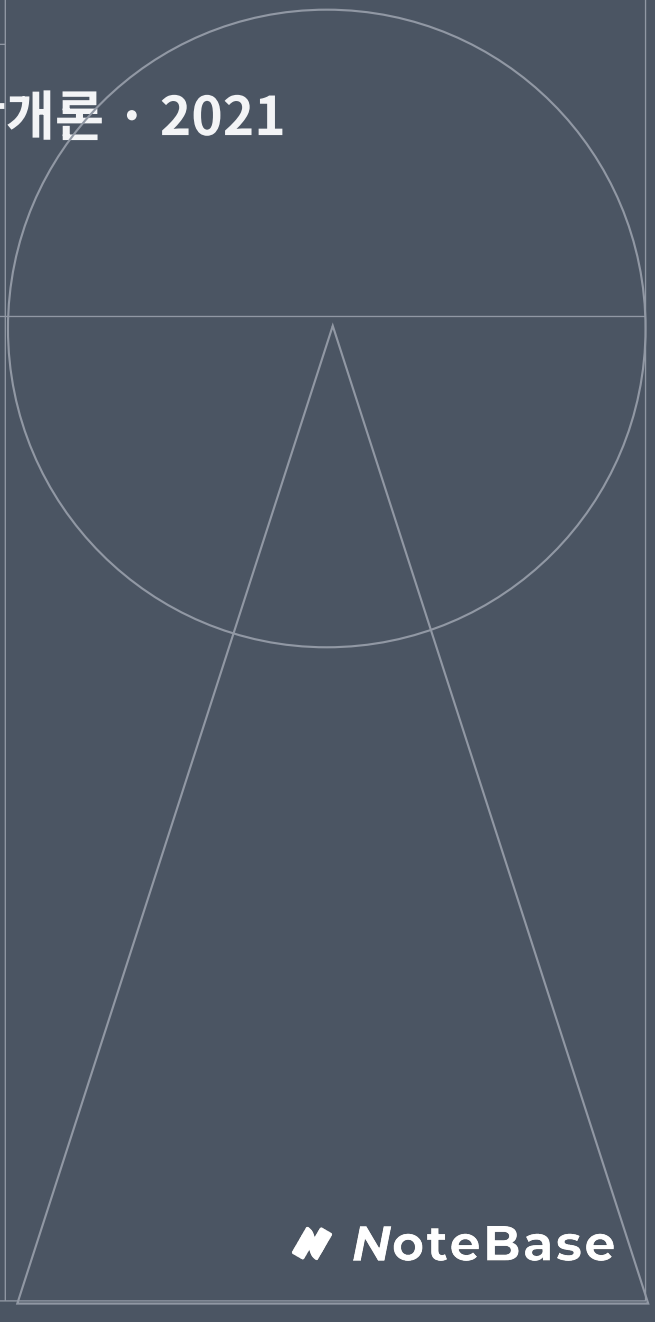
기출OX 해설집

공무원/국회직/9급 · 정보학개론 · 2021



NoteBase

기출OX 해설집

공무원/국회직/9급 · 정보학개론 · 2021

공무원/국회직/9급 · 정보학개론 · 2021

1. 축적 검색(Ad hoc retrieval)에 대한 설명으로 <보기>에서 옳은 것만을 모두 고르면?

<보 기>

ㄱ. 비교적 정적인(static) 정보원을 대상으로 새로운 질의를 입력한다.
ㄴ. 새로 입력된 정보자료를 대상으로 하여 검색하는 작업이다.
ㄷ. 소급 검색에 상응한다.
ㄹ. 선택적 정보제공 서비스에 상응한다.

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄱ, ㄹ
④ ㄴ, ㄷ
⑤ ㄴ, ㄹ

정답 2번

출제 포인트

축적 검색(Ad hoc retrieval)의 검색 대상과 질의 방식, 그리고 소급 검색·선택적 정보제공 서비스와의 차이를 묻는 문제이다.

핵심 개념

축적 검색: 비교적 정적인 정보원에 이용자가 그때그때 새로운 질의를 입력하여 이미 축적된 자료를 검색하는 방식이다. / 결정 차이축: 정보원은 기존에 축적되어 있고 질의가 새로 들어온다. / 키워드: 정적 정보원·새 질의·소급 검색

선택적 정보제공(SDI): 미리 등록한 이용자 관심 프로파일에 새로 들어온 정보를 계속 대조하여 선별 제공하는 서비스이다. / 결정 차이축: 질의·프로파일은 비교적 지속되고 새로운 정보가 들어온다. / 키워드: 새 정보·지속적 선별 제공

따라서 축적 검색은 '정적 정보원+새 질의', 선택적 정보제공은 '지속 질의+새 정보'로 대비된다.

NoteBase

선지별 해설

① X — ㄱ은 맞지만 ㄴ은 틀리고, 옳은 ㄷ이 빠져 있다. '새로 입력된 정보자료를 대상으로 검색한다'는 ㄴ은 선택적 정보제공에 가까운 설명이며, 올바른 조합은 ㄱ, ㄷ이다.
② O — ㄱ의 '비교적 정적인 정보원에 새로운 질의를 입력한다'와 ㄷ의 '소급 검색에 상응한다'가 모두 축적 검색의 특성에 해당한다.
③ X — ㄹ의 '선택적 정보제공 서비스에 상응한다'가 틀리고, 옳은 ㄷ이 빠져 있다. 축적 검색은 선택적 정보제공이 아니라 소급 검색에 상응하므로 올바른 조합은 ㄱ, ㄷ이다.
④ X — ㄷ은 맞지만 ㄴ이 틀리고, 옳은 ㄱ이 빠져 있다. 축적 검색은 새로 입력된 정보자료가 아니라 비교적 정적인 축적 정보원을 새 질의로 검색하므로 올바른 조합은 ㄱ, ㄷ이다.
⑤ X — ㄴ과 ㄹ이 모두 틀리고, 옳은 ㄱ과 ㄷ이 빠져 있다. 새 정보자료를 지속적으로 선별 제공하는 것은 선택적 정보제공의 성격이며, 축적 검색의 올바른 조합은 ㄱ, ㄷ이다.

오답 포인트

'새로운 질의'와 '새로 입력된 정보자료'를 바꾸어 읽기 쉬운 점이 함정이다. 축적 검색에서 새롭게 제시되는 것은 질의이고, 선택적 정보제공에서 계속 새로 들어오는 것은 정보자료이다.

다시 나오면 이렇게 판단

무엇이 새로 들어오는지를 기준으로 구별한다. 기존 축적 자료에 새 질의가 들어오면 축적 검색·소급 검색이고, 기존 관심 프로파일에 새 정보가 들어오면 선택적 정보제공이다.

암기 포인트

축적 검색은 '자료는 축적, 질의는 새로', 선택적 정보제공은 '질의는 지속, 정보는 새로'이다.

NoteBase

2. 시소러스에 대한 설명으로 옳지 않은 것은?

① 색인과 검색에서 표준화된 어휘를 제공한다.
② 자연어로 된 용어들의 통제어휘로 후조합을 위해 설계되었다.
③ 분석 방식으로 열거 방식을 취하며 유연성이 적다.
④ 용어들 간의 동등관계, 계층관계, 연관관계를 표현할 수 있다.
⑤ 색인어휘와 도입어휘로 구성되며, 색인어휘는 우선어, 도입어휘는 비우선어라고 한다.

정답 3번

출제 포인트

시소러스의 기능, 용어 관계, 구성 요소와 후조합 방식의 특성을 구별하는 문제이다.

핵심 개념

시소러스는 색인과 검색에 쓰는 용어를 통제하고, 용어 사이의 의미 관계를 체계적으로 나타낸 통제어휘집이다. 자연어 표현을 우선어인 디스크립터로 통일하며, 비우선어인 비디스크립터는 우선어로 안내한다. 용어 관계에는 동등관계, 계층관계, 연관관계가 있으며, 여러 용어를 필요에 따라 결합하는 후조합 방식이므로 유연성이 크다.

선지별 해설

① O — 시소러스는 동의어·유사어 등 다양한 표현을 표준화된 통제어휘로 연결하여 색인어와 검색어를 일치시키는 데 도움을 준다.
② O — 자연어로 표현된 여러 용어를 통제어휘로 정리하여 동의어와 표현 차이로 생기는 검색 누락을 줄이도록 설계된다.
③ X — 틀린 표현은 '열거 방식을 취하며 유연성이 적다'이다. 시소러스는 개념을 나타내는 용어들을 분석하고 검색 시 필요한 용어를 결합하는 후조합 방식을 취하므로 유연성이 크다고 고쳐야 한다.
④ O — 시소러스는 동의어나 이형 표현을 잇는 동등관계, 상위어와 하위어를 잇는 계층관계, 의미상 관련된 용어를 잇는 연관관계를 표현할 수 있다.
⑤ O — 시소러스는 색인과 검색에 실제로 사용하는 우선어인 디스크립터와, 그 우선어로 안내하는 비우선어인 비디스크립터로 구성된다. 따라서 색인어휘는 우선어, 도입어휘는 비우선어에 해당한다.

오답 포인트

③은 시소러스에 용어가 미리 수록된다는 점을 '열거식 분류'와 혼동하게 만든다. 시소러스의 핵심은 용어를 필요한 시점에 결합하는 후조합성과 그에 따른 높은 유연성이다.

NoteBase

다시 나오면 이렇게 판단

시소러스 선지는 '통제어휘 제공', '우선어와 비우선어', '동등·계층·연관관계', '후조합과 높은 유연성'의 네 기준으로 판단한다. 이 기준과 반대로 열거 방식이나 낮은 유연성을 제시하면 틀린 설명이다.

암기 포인트

시소러스는 용어를 통제하고 관계를 연결하며, 후조합으로 유연하게 검색한다.

NoteBase

3. 자동색인에 대한 설명으로 옳지 않은 것은?

① 문헌의 본문에 나타나는 단어나 단어구를 컴퓨터로 추출하여 작성한다.

② 통계적 기법은 문헌에 나타난 단어의 출현빈도를 이용하며, 비교적 구현작업이 간단하다.

③ 확률색인 기법은 문헌 내에서 단어의 관계를 확률에 기초하여 색인하는 것으로 통계적·비통계적 기법과 다른 별도의 기법이다.

④ 단어가중 기법에서 TF*IDF 기법은 단어빈도와 문헌빈도를 사용하여 계산한다.

⑤ 한글문헌의 자동색인에서 색인어 추출은 어절이 곧 단어가 되지 않으므로 어절 분리 후 이차적인 처리가 필요하다.

정답 3번

출제 포인트

자동색인의 기본 과정과 통계적 색인, 확률색인, TF-IDF, 한국어 형태소 처리의 관계를 구별하는 문제이다.

핵심 개념

자동색인은 문헌의 본문에서 색인어 후보를 컴퓨터로 추출하고 가중치를 계산하여 색인어를 선정하는 방식이다. 통계적 색인은 단어의 출현빈도와 분포 같은 통계량을 이용한다. 확률색인은 문헌과 단어의 관계를 확률로 표현하는 통계적 색인 기법에 속한다.
TF-IDF는 문헌 안에서 자주 나오는 단어에 높은 값을 주는 단어빈도(TF)와, 여러 문헌에 널리 나오는 단어의 값을 낮추는 역문헌빈도(IDF)를 결합한 가중치이다.

선지별 해설

① O — 자동색인은 문헌 본문에 나타난 단어나 단어구를 컴퓨터로 추출하여 색인어를 작성하는 방식이므로 옳다.
② O — 통계적 기법은 단어의 출현빈도와 분포를 주요 근거로 삼으며, 빈도 계수와 가중치 계산을 중심으로 하므로 비교적 구현이 간단하다.
③ X — 틀린 표현은 '통계적·비통계적 기법과 다른 별도의 기법'이다. 확률색인 기법은 문헌과 단어의 관계를 확률에 기초하여 판단하는 통계적 기법의 한 유형이다.
④ O — TF-IDF는 단어빈도(TF)와 문헌빈도에 역으로 비례하도록 산출한 역문헌빈도(IDF)를 결합하여 계산한다. 따라서 단어빈도와 문헌빈도를 이용한다는 설명은 옳다.
⑤ O — 한국어의 어절은 조사나 어미 등이 결합된 형태여서 어절 자체가 곧 색인어가 되지 않을 수 있다. 따라서 어절 분리 뒤 형태소 분석 등 추가 처리를 거쳐 색인어 후보를 추출해야 한다.

NoteBase

오답 포인트

③은 '확률'이라는 명칭 때문에 통계적 기법과 독립된 방식처럼 보이게 만든 선지이다. 그러나 확률은 문헌과 단어의 관련성을 수량화하는 통계적 판단 근거이므로 확률색인법은 통계적 색인 범주에 포함된다. 또한 ④의 문헌빈도는 그대로 가중치를 높이는 것이 아니라 IDF에서 역방향으로 반영된다는 점을 함께 구별해야 한다.

다시 나오면 이렇게 판단

선지의 분류 관계를 먼저 확인한다. 출현빈도·문헌빈도·확률을 수치로 이용하면 통계적 색인이고, TF-IDF에서는 한 문헌 내 빈도인 TF와 문헌 집합 전체의 희소성을 나타내는 IDF를 구별한다. 한국어 자동색인은 어절 추출 뒤 형태소 분석 등 후속 처리가 필요하다고 판단한다.

암기 포인트

확률색인은 통계적 색인에 포함되고, TF-IDF는 문헌 안에서는 자주 나오되 문헌 전체에서는 드문 단어를 높게 평가한다.

NoteBase

4. <보기>에서 설명하는 것은?

<보 기>

○ SF 작가 닐 스티븐슨(Neal Stephenson)의 1992년작『스노 크래시(Snow Crash)』에 묘사된 가상세계 메타버스를 3차원 컴퓨터 그래픽 환경으로 구현해 보자는 취지 아래 개발하였다.

○ 그리드 컴퓨팅기술과 CGS라는 미디어 객체 창작 자동화 기술을 최적화하여 가상세계 내에서 사용자들이 원하는 대로 공간을 확장해 가고, 자신의 창의성을 실현할 수 있도록 하였다.

① Second Life
② Library Park
③ Maker Space
④ Digital Larchiveum
⑤ Cloud Digital Library

정답 1번

출제 포인트

이 문제는 소설 『스노 크래시』의 메타버스 개념을 3차원 컴퓨터 그래픽으로 구현하고, 이용자가 가상공간을 직접 확장하도록 한 서비스가 무엇인지 묻는다.

핵심 개념

Second Life는 린든 랩이 개발한 3차원 가상세계 서비스로, 이용자가 아바타로 활동하며 가상공간과 콘텐츠를 만들고 확장할 수 있다.
Library Park는 도서관과 공원적 기능을 결합한 복합 문화·휴식 공간이고, Maker Space는 이용자가 장비와 공간을 공유하며 직접 제작·실험하는 창작 공간이다.
Digital Larchiveum은 Library·Archives·Museum의 기능과 자료를 디지털 환경에서 연계한 개념이며, Cloud Digital Library는 클라우드 컴퓨팅을 활용하여 디지털 자료와 서비스를 제공하는 도서관이다.

NoteBase

선지별 해설

① O — Second Life는 린든 랩이 구현한 3차원 가상세계이며, 이용자가 원하는 공간과 콘텐츠를 만들어 가상세계를 확장한다는 보기의 설명과 일치한다.
② X — 틀린 표현은 ‘Library Park’이다. Library Park는 도서관과 공원적 기능을 결합한 공간을 가리키며, 보기의 가상세계 서비스 명칭은 ‘Second Life’로 고쳐야 한다.
③ X — 틀린 표현은 ‘Maker Space’이다. Maker Space는 공동 장비와 작업 공간을 이용해 창작물을 직접 제작하는 공간이며, 보기의 개념은 ‘Second Life’로 고쳐야 한다.
④ X — 틀린 표현은 ‘Digital Larchiveum’이다. Digital Larchiveum은 도서관·기록관·박물관의 디지털 자원과 기능을 연계한 개념이며, 보기의 개념은 ‘Second Life’로 고쳐야 한다.
⑤ X — 틀린 표현은 ‘Cloud Digital Library’이다. Cloud Digital Library는 클라우드 컴퓨팅 기반으로 디지털 자료와 도서관 서비스를 제공하는 체계이며, 보기의 개념은 ‘Second Life’로 고쳐야 한다.

오답 포인트

‘그리드 컴퓨팅’, ‘미디어 객체 창작 자동화’ 같은 기술 용어만 보고 디지털 도서관 계열인 ④나 ⑤를 고르기 쉽다. 그러나 『스노 크래시』의 메타버스, 린든 랩, 이용자의 가상공간 창작이라는 단서를 함께 만족하는 것은 Second Life이다.

다시 나오면 이렇게 판단

‘린든 랩 + 3차원 가상세계 + 이용자 제작 콘텐츠’가 제시되면 Second Life로 판단한다. 자료를 보존·연계·제공하는 도서관 개념과 이용자가 활동하고 세계를 만들어 가는 가상세계 서비스를 구분한다.

암기 포인트

린든 랩의 이용자 창작형 3차원 가상세계는 Second Life이다.

NoteBase

5. 토픽맵(Topic Map)에 대한 설명으로 옳지 않은 것은?

① 토픽맵은 정보들을 토픽으로 표현하고, 표현된 토픽들 간의 주제적 연계성을 관계로 단순화하여 기계가 이해할 수 있도록 제공한다.

② 토픽은 주제가 되는 하나의 지식 단위를 표현하는 것으로, 토픽의 생성은 기계가 이해하고 처리할 수 있는 하나의 객체가 생성되는 것을 말한다.

③ 정보자원들의 다차원적인 속성에 의해 상호 연관성 있게 연결하여 지식 구조를 표현한다.

④ 하나의 관계에 연결될 수 있는 토픽의 수는 여러 개가 될 수 있으며, 토픽들 간의 관계 또한 토픽의 수에 상관없이 무한하게 설정할 수 있다.

⑤ 관계(association)는 토픽과 관련된 하나 이상의 정보자원을 의미하며, 이 정보자원은 토픽과 연결된다.

정답 5번

출제 포인트

토픽맵을 구성하는 토픽(topic), 관계(association), 발생(occurrence)의 역할을 구별하는 문제이다.

핵심 개념

토픽(topic): 어떤 주제를 기계가 처리할 수 있도록 표현한 지식 단위이다.
관계(association): 둘 이상의 토픽이 서로 어떤 의미 관계를 맺는지를 표현한다.
발생(occurrence): 특정 토픽과 그 토픽에 관한 정보자원을 연결한다.
핵심 대비는 '토픽과 토픽의 연결은 관계', '토픽과 정보자원의 연결은 발생'이다.

선지별 해설

① O — 토픽맵은 정보를 토픽으로 표현하고 토픽 사이의 주제적 연관성을 관계로 구조화하여 기계가 처리할 수 있게 한다.
② O — 토픽은 주제가 되는 대상을 정보 시스템에서 식별하고 처리할 수 있도록 나타낸 객체이다.
③ O — 토픽맵은 정보자원의 여러 속성과 연관성을 토픽 중심으로 연결하여 지식 구조를 표현한다.
④ O — 하나의 관계에는 둘 이상의 여러 토픽이 참여할 수 있으며, 토픽들 사이에 필요한 관계를 여러 개 설정할 수 있다.
⑤ X — 틀린 표현은 '관계(association)는 토픽과 관련된 하나 이상의 정보자원을 의미하며, 이 정보자원은 토픽과 연결된다'이다. 관계는 토픽들 사이의 의미적 연결을 나타낸다. 올바르게는 토픽과 관련 정보자원을 연결하는 것은 발생(occurrence)이다.

NoteBase

오답 포인트

⑤는 '연결'이라는 공통 표현을 이용해 관계와 발생을 바꾸어 놓았다. 연결 대상이 토픽들인지, 토픽과 정보자원인지가 판단의 핵심이다.

다시 나오면 이렇게 판단

연결되는 양쪽 대상을 먼저 확인한다. 토픽과 토픽을 잇는다면 관계(association), 토픽과 관련 정보자원을 잇는다면 발생(occurrence)으로 판단한다.

암기 포인트

토픽끼리는 관계(association), 토픽과 자료는 발생(occurrence)이다.

NoteBase

6. 다음은 A 학술지의 2020년 IF(Impact Factor) 계산식이다. 이에 대한 설명으로 옳은 것은?

$$\text{A 학술지의 2020년 IF} = \frac{326}{219} = 1.489$$

① IF 공식의 분자는 A 학술지에 수록된 논문의 수이다.
② IF 공식의 분모는 A 학술지의 논문이 인용된 총 횟수이다.
③ 2020년 IF 값을 계산하기 위해 2018, 2019년에 출판된 논문을 이용한다.
④ IF 값은 학술지의 영향력을 평가하므로 1.0 이하를 가질 수 없다.
⑤ IF는 미국의 Institute for Scientific Information 회사를 설립한 룬(Luhn)에 의해 고안되었다.

정답 3번

출제 포인트

이 문제는 특정 연도의 학술지 영향력지수(IF)를 계산할 때 분자와 분모에 무엇이 들어가는지, 그리고 어느 연도에 출판된 논문을 대상으로 하는지를 구분하는 문제이다.

조건 정리

주어진 값: 2020년에 받은 인용 횟수는 326회이고, 직전 2년인 2018년과 2019년에 출판된 인용 가능 논문 수는 219편이다.
구해야 할 값: A 학술지의 2020년 IF와 각 선지의 옳고 그름이다.
주의할 조건: 평가연도인 2020년에 출판된 논문 수를 쓰는 것이 아니라 직전 2개 연도에 출판된 논문을 기준으로 한다.

사용할 공식

2020년 IF=2018~2019년 출판 논문이 2020년에 받은 인용 횟수2018~2019년 출판된 인용 가능 논문 수
이 공식은 한 편의 논문이 평균적으로 얼마나 인용되었는지를 나타내기 때문에, 인용 횟수를 논문 수로 나눈다.

계산 과정

1. 분자는 2018년과 2019년에 출판된 대상 논문이 2020년에 받은 인용 횟수 326회이다.
2. 분모는 2018년과 2019년에 출판된 인용 가능 논문 수 219편이다.
3. 따라서 $\frac{326}{219} \approx 1.489$이므로 2020년 IF는 1.489이다.

NoteBase

선지 판단

① X — 분자를 단순히 'A 학술지에 수록된 논문의 수'라고 한 부분이 틀리다. 올바르게는 직전 2년인 2018년과 2019년에 출판된 대상 논문이 2020년에 받은 인용 횟수이다.
② X — 분모를 'A 학술지의 논문이 인용된 총 횟수'라고 한 부분이 틀리다. 올바르게는 2018년과 2019년에 A 학술지에 출판된 인용 가능 논문 수이다.
③ O — 2020년 IF는 평가연도의 직전 2개 연도인 2018년과 2019년에 출판된 논문을 대상으로 계산한다.
④ X — '1.0 이하를 가질 수 없다'는 부분이 틀리다. IF는 평균 인용 횟수이므로 1.0 이하가 될 수도 있고 0이 될 수도 있다.
⑤ X — IF를 룬(Luhn)이 고안했다는 부분이 틀리다. 올바르게는 유진 가필드(Eugene Garfield)가 IF를 고안하고 Institute for Scientific Information을 설립하였다.

오답 포인트

핵심 함정은 분자와 분모를 뒤바꾸거나, 평가연도와 논문 출판연도를 혼동하는 것이다. 분자는 인용 횟수, 분모는 논문 수이며, 2020년 IF에서는 2020년 논문이 아니라 2018년과 2019년 논문을 대상으로 한다. 또한 IF는 평균값이므로 반드시 1.0을 넘는다는 하한이 없다.

다시 나오면 이렇게 판단

IF 문제에서는 '평가연도에 받은 인용 횟수 ÷ 직전 2년간 출판된 인용 가능 논문 수'의 구조를 먼저 확인한다. 이어서 분자는 횟수, 분모는 편수인지 검토하고 평가연도에서 1년과 2년을 뺀 출판연도를 대입한다.

암기 포인트

IF는 '올해 받은 인용 ÷ 직전 2년 논문 수'이며, 고안자는 유진 가필드이다.

NoteBase

7. 통제어 색인에 대한 설명으로 옳지 않은 것은?

① 색인어 선정 시, 통제된 색인어휘집에서 색인어를 할당한다.
② 정보자료에서 색인어를 발췌한다고 하여 발췌색인(extraction index)이라고 한다.
③ 특정한 개념은 항상 같은 용어에 의해 색인이 가능하므로, 검색 시 이 용어를 검색어로 사용하면 관련된 모든 정보자료를 검색할 수 있다.
④ 주제명표목표와 시소러스는 통제어 색인을 위해 사용되는 가장 대표적인 색인도구이다.
⑤ 색인될 개념이 새로운 것이거나, 고유명사이거나, 통제어휘에서 해당 용어를 찾을 수 없는 경우에는 식별어(identifier)가 만들어질 수 있다.

정답 2번

출제 포인트

통제어 색인과 자연어·발췌색인을 색인어의 선정 출처에 따라 구별하고, 통제어휘집의 기능과 예외 처리 방법을 이해하는 문제이다.

핵심 개념

통제어 색인은 문헌의 개념을 분석한 뒤 통제어휘집에 미리 수록된 표준 용어를 색인어로 할당하는 방식이다.
발췌색인은 정보자료에 실제로 나타난 자연어의 단어나 문구를 발췌하여 색인어로 채택하는 방식이다.
통제어휘집은 동의어를 하나의 우선어로 모으고 용어 사이의 계층·연관 관계를 제시하여 색인과 검색의 일관성을 높인다.

선지별 해설

① O — 통제어 색인에서는 문헌의 개념을 분석한 후 통제된 색인어휘집에서 적절한 색인어를 골라 할당한다.
② X — 틀린 표현은 '정보자료에서 색인어를 발췌한다'이다. 정보자료 안에 완성된 색인어가 미리 존재하는 것이 아니라, 자료에 나타난 자연어의 단어나 문구를 발췌하여 색인어로 선정하는 방식을 발췌색인이라고 한다.
③ O — 어휘통제를 하면 같은 개념을 동일한 우선어로 일관되게 색인할 수 있으므로, 그 우선어를 검색어로 사용하여 관련 자료를 모아 검색할 수 있다.
④ O — 주제명표목표와 시소러스는 표준화된 색인어와 그 관계를 제공하는 대표적인 통제어 색인도구이다.
⑤ O — 새 개념, 고유명사 등으로 통제어휘집에서 대응 용어를 찾을 수 없을 때에는 그 대상을 직접 나타내는 식별어를 새로 만들어 사용할 수 있다.

NoteBase

오답 포인트

②는 '발췌색인'이라는 명칭 자체가 아니라 발췌 대상을 바꾸어 놓은 선지이다. 자료에서 발췌하는 것은 이미 확정된 색인어가 아니라 자연어로 표현된 단어나 문구이며, 이를 선별한 결과가 색인어가 된다.

다시 나오면 이렇게 판단

색인어를 어디에서 얻는지를 확인한다. 통제어휘집에서 표준 용어를 골라 붙이면 통제어·할당색인이고, 원문에 나온 자연어의 단어나 문구를 뽑아 색인어로 삼으면 자연어·발췌색인이다.

암기 포인트

통제어휘집에서 골라 붙이면 할당, 원문의 자연어를 뽑아 쓰면 발췌이다.

NoteBase

8. <보기>의 내용을 바탕으로 올바른 검색을 수행한 것은?

<보 기>

'program'이란 단어는 적어도 세 가지 다른 개념을 표현하는 데 사용된다. 이 중 하나의 개념에 대해 탐색하고자 한다.

① 검색의 확대를 위해 인접 탐색 연산자를 사용하였다.
② 관련이 없는 문헌을 제거하기 위해 퍼지 탐색 기법을 사용하였다.
③ 검색의 축소를 위해 검색어 computer와 불리언 연산자 OR를 적용하였다.
④ 적합 순위별 검색 결과를 얻기 위해 관련 용어와 함께 가중치 탐색 기법을 사용하였다.
⑤ 검색의 정확률을 높이기 위해 program의 세 가지 개념에 대한 동의어들을 찾아서 검색어로 추가하였다.

정답 4번

출제 포인트

이 문제는 여러 의미를 가진 검색어에서 특정 개념에 관한 문헌을 찾을 때 검색 확대·축소·정교화·순위화 기법의 목적을 구별하는 문제이다.

핵심 개념

인접 탐색은 둘 이상의 검색어가 서로 가까이 나타나도록 제한하여 검색 결과를 정교화하는 기법이다. 패싯 탐색은 주제·자료 유형·연도 등 분류 속성을 선택하여 결과 범위를 단계적으로 좁히는 기법이다. 불리언 OR는 어느 하나의 검색어만 포함해도 결과에 넣으므로 검색 범위와 재현율을 높인다. 가중치 탐색은 특정 검색어나 개념에 더 큰 중요도를 부여하여 적합도가 높은 문헌이 상위에 나타나게 하는 기법이다.

NoteBase

선지별 해설

① X — '검색의 확대를 위해 인접 탐색 연산자를 사용하였다'는 부분이 틀렸다. 인접 탐색은 검색어 사이의 거리나 순서를 제한하여 결과를 정교화하거나 줄이는 데 사용한다. 검색을 확대하려면 동의어·관련어를 OR로 결합하는 방법 등을 사용한다.
② X — '관련이 없는 문헌을 제거하기 위해 패싯 탐색 기법을 사용하였다'는 표현은 부정확하다. 패싯 탐색은 문헌의 관련성 자체를 자동 판정하여 제거하는 기법이 아니라, 이용자가 원하는 주제·형식·연도 등의 패싯 값을 선택하여 결과 범위를 좁히는 기법이다. 올바르게는 '원하는 패싯 값을 선택하여 검색 결과를 단계적으로 축소하였다'라고 해야 한다.
③ X — '검색의 축소를 위해 검색어 computer와 불리언 연산자 OR를 적용하였다'는 부분이 틀렸다. OR는 program 또는 computer 중 하나만 포함한 문헌도 검색하므로 결과를 확대한다. 검색을 축소하려면 AND, NOT, 인접 연산자 또는 구체적인 제한 조건을 사용한다.
④ O — 탐색하려는 개념과 관련된 용어에 가중치를 주면 그 개념과 밀접한 문헌의 적합도 점수가 높아져 검색 결과를 적합도 순으로 제시할 수 있다. 여러 의미를 가진 program 중 하나의 개념을 집중적으로 탐색하려는 상황에 알맞다.
⑤ X — '정확률을 높이기 위해 세 가지 개념의 동의어들을 모두 검색어로 추가하였다'는 부분이 틀렸다. 세 개념의 동의어를 모두 추가하면 검색 범위와 재현율은 커질 수 있지만, 원하지 않는 다른 의미의 문헌까지 늘어 정확률이 낮아질 수 있다. 정확률을 높이려면 탐색하려는 한 개념의 관련어만 선택하고 가중치나 제한 조건을 적용해야 한다.

오답 포인트

③의 OR를 검색 축소 연산자로 착각하거나, ⑤에서 동의어를 많이 넣으면 언제나 정확률이 높아진다고 생각하기 쉽다. OR와 동의어 추가는 일반적으로 검색 결과를 늘려 재현율을 높이며, 특정 의미의 문헌을 위로 올리는 기능은 가중치 부여와 적합도 순위화에 해당한다.

다시 나오면 이렇게 판단

검색 기법을 목적별로 구분한다. OR·동의어 추가는 확대, AND·NOT·인접 조건·패싯 제한은 축소 또는 정교화, 가중치 부여는 적합도 순위화로 판단한다. 다의어에서 특정 의미를 우선하려는 경우에는 그 의미의 관련어에 가중치를 주었는지 확인한다.

암기 포인트

OR는 넓게, 인접·패싯 제한은 좁게, 가중치는 적합한 문헌을 위로 올린다.

NoteBase

9. 한글 자연언어 처리 과정을 순서대로 바르게 나열한 것은?

① 형태소 분석 → 의미 분석 → 구문 분석 → 화용 분석
② 형태소 분석 → 구문 분석 → 의미 분석 → 화용 분석
③ 구문 분석 → 형태소 분석 → 의미 분석 → 화용 분석
④ 구문 분석 → 형태소 분석 → 화용 분석 → 의미 분석
⑤ 의미 분석 → 형태소 분석 → 구문 분석 → 화용 분석

정답 2번

출제 포인트

한글 자연언어 처리의 각 분석 단계가 어떤 정보를 입력으로 삼는지 이해하고 올바른 처리 순서를 판단하는 문제이다.

올바른 순서

1. 형태소 분석 — 문장을 의미를 지닌 최소 단위인 형태소로 나누고 품사 등의 정보를 분석한다.
2. 구문 분석 — 형태소 분석 결과를 바탕으로 문장 성분 사이의 문법적 구조와 의존 관계를 파악한다.
3. 의미 분석 — 구문 구조를 바탕으로 단어와 문장이 나타내는 의미를 해석한다.
4. 화용 분석 — 발화 상황, 문맥, 화자의 의도 등을 고려하여 실제 전달 의미를 파악한다.

왜 이 순서인가: 앞 단계에서 얻은 형태와 구조 정보가 뒤 단계의 의미 및 문맥 해석을 위한 입력이 되기 때문이다.

선지별 해설

① X — 의미 분석을 구문 분석보다 먼저 둔 부분이 틀렸다. 문장의 구조를 파악한 뒤 그 구조에 따라 의미를 해석한다. 올바르게는 형태소 분석 → 구문 분석 → 의미 분석 → 화용 분석이다.
② O — 형태소 단위 분석, 문법 구조 분석, 문장 의미 해석, 문맥과 의도 해석의 순서로 올바르게 배열하였다.
③ X — 구문 분석을 형태소 분석보다 먼저 둔 부분이 틀렸다. 구문 분석은 형태소 분석에서 얻은 단위와 품사 정보를 바탕으로 이루어진다. 올바르게는 형태소 분석 → 구문 분석 → 의미 분석 → 화용 분석이다.
④ X — 구문 분석을 형태소 분석보다 먼저 둔 데다 화용 분석을 의미 분석보다 먼저 둔 부분도 틀렸다. 문장 구조와 의미를 파악한 뒤 문맥에 따른 실제 의미를 해석한다. 올바르게는 형태소 분석 → 구문 분석 → 의미 분석 → 화용 분석이다.
⑤ X — 의미 분석을 형태소 분석과 구문 분석보다 먼저 둔 부분이 틀렸다. 의미 분석에는 앞서 파악한 형태소 정보와 구문 구조가 필요하다. 올바르게는 형태소 분석 → 구문 분석 → 의미 분석 → 화용 분석이다.

오답 포인트

형태소 분석과 구문 분석의 선후관계, 의미 분석과 화용 분석의 차이가 핵심 함정이다. 구문 분석은 형태소 분석 결과에 의존하며, 화용 분석은 문장 자체의 의미를 넘어 문맥과 화자의 의도까지 해석하므로 의미 분석 뒤에 놓인다.

NoteBase

다시 나오면 이렇게 판단

처리 범위가 작은 언어 단위에서 큰 해석 단위로 확장되는 흐름을 적용한다. 단어의 구성과 품사인 형태를 먼저 보고, 문장 구조인 구문을 파악한 다음, 문장 자체의 의미를 해석하고, 마지막으로 상황과 의도를 포함한 화용을 판단한다.

암기 포인트

형태 → 구조 → 의미 → 맥락, 즉 형태소 → 구문 → 의미 → 화용의 순서이다.

NoteBase

10. 검색시스템에서 다음 <조건>에 맞는 재현율과 정확률을 바르게 계산한 것은?

<조 건>

○ 논문의 총 수는 1,000건
○ 전체 논문에서 '정보검색' 관련 적합논문은 300건
○ 홍길동이 '정보검색'으로 검색한 결과는 200건이고, 이 중에서 부적합한 논문은 50건

① 재현율 = 50%, 정확률 = 75%
② 재현율 = 75%, 정확률 = 50%
③ 재현율 = 30%, 정확률 = 20%
④ 재현율 = 20%, 정확률 = 75%
⑤ 재현율 = 50%, 정확률 = 50%

정답 1번

출제 포인트

이 문제는 검색된 적합 문헌 수를 먼저 구한 뒤 재현율과 정확률의 분모를 구분하여 계산하는 문제이다.

조건 정리

전체 논문 수: 1,000건
전체 적합 문헌 수: 300건
검색된 문헌 수: 200건
검색된 부적합 문헌 수: 50건
검색된 적합 문헌 수: 200건 - 50건 = 150건
구해야 할 값: 재현율과 정확률

사용할 공식

재현율=검색된 적합 문헌 수전체 적합 문헌 수× 100
정확률=검색된 적합 문헌 수검색된 전체 문헌 수× 100
재현율은 전체 적합 문헌을 얼마나 빠짐없이 검색했는지를 나타내므로 전체 적합 문헌 수가 분모가 된다.
정확률은 검색 결과 중 적합 문헌이 차지하는 비율이므로 검색된 전체 문헌 수가 분모가 된다.

NoteBase

계산 과정

1. 검색된 적합 문헌 수는 200건 - 50건 = 150건이다.
2. 재현율은 $\frac{150}{300} \times 100 = 50\%$이다.
3. 정확률은 $\frac{150}{200} \times 100 = 75\%$이다.

선지 판단

① O — 재현율 50%, 정확률 75%로 계산 결과와 모두 일치한다.
② X — '재현율 75%, 정확률 50%'는 두 비율을 서로 바꾼 값이다. 올바르게는 재현율 50%, 정확률 75%이다.
③ X — '재현율 30%, 정확률 20%'는 전체 논문 1,000건을 각각의 분모처럼 잘못 사용한 값이다. 올바르게는 재현율의 분모가 300건, 정확률의 분모가 200건이므로 재현율 50%, 정확률 75%이다.
④ X — '재현율 20%'는 검색된 전체 문헌 200건을 전체 논문 1,000건으로 나눈 값으로, 재현율 공식에 맞지 않는다. 올바르게는 재현율이 150/300 = 50%이고 정확률이 75%이다.
⑤ X — '정확률 50%'는 재현율 값을 정확률에도 잘못 적용한 것이다. 올바르게는 정확률이 150/200 = 75%이고 재현율이 50%이다.
따라서 재현율 50%, 정확률 75%이므로 정답은 1번이다.

오답 포인트

핵심 함정은 재현율과 정확률의 분모를 뒤바꾸는 것이다. 또한 검색된 200건에는 부적합 문헌 50건이 포함되므로, 두 공식의 분자에는 200건이 아니라 검색된 적합 문헌 150건을 넣어야 한다.

다시 나오면 이렇게 판단

먼저 검색된 적합 문헌 수를 '검색된 전체 문헌 수 - 검색된 부적합 문헌 수'로 구한다. 그다음 재현율은 전체 적합 문헌을 분모로, 정확률은 검색된 전체 문헌을 분모로 하여 계산한다.

암기 포인트

재현율은 적합 문헌을 얼마나 빠짐없이 찾았는지, 정확률은 검색 결과가 얼마나 정확한지 나타낸다. 재현율의 분모는 전체 적합 문헌, 정확률의 분모는 검색된 전체 문헌이다.

NoteBase

11. 정보검색의 벡터공간 모델에 대한 설명으로 옳지 않은 것은?

① 용어의 벡터를 기반으로 한다.
② 문헌과 질의에 나타난 용어들 간 부분 매칭이 가능하다.
③ 검색 결과의 순위화가 가능하다.
④ 동의어나 어구의 관계를 나타내는 데 용이하다.
⑤ 수학적으로는 대수론(algebra theory)에 근거한다.

정답 4번

출제 포인트

이 문제는 정보검색의 벡터공간 모델이 용어 벡터와 유사도를 이용하는 방식임을 이해하고, 그 장점과 한계를 구별하는 문제이다.

핵심 개념

벡터공간 모델은 문서와 질의를 용어별 가중치로 이루어진 벡터로 표현하고, 두 벡터의 유사도를 계산하여 검색 결과의 순위를 정하는 대수적 정보검색 모델이다.
부분 매칭: 질의 용어가 문서에 일부만 포함되어도 유사도를 계산할 수 있다.
의미 관계의 한계: 기본 모델은 각 용어를 서로 독립된 차원으로 취급하므로 동의어나 여러 단어로 이루어진 어구의 의미 관계를 직접 나타내기 어렵다.

선지별 해설

① O — 문서와 질의를 용어의 가중치로 구성된 벡터로 표현하므로 용어 벡터를 기반으로 한다.
② O — 문서 벡터와 질의 벡터 사이의 유사도를 계산하므로 질의 용어가 문서에 일부만 나타나는 경우에도 부분 매칭이 가능하다.
③ O — 계산된 유사도 점수에 따라 문서를 정렬할 수 있으므로 검색 결과의 순위화가 가능하다.
④ X — 틀린 표현은 '동의어나 어구의 관계를 나타내는 데 용이하다'이다. 기본적인 벡터공간 모델은 용어를 독립된 차원으로 다루므로 동의어나 어구의 의미 관계를 직접 나타내기 어렵다고 고쳐야 한다.
⑤ O — 문서와 질의를 벡터로 나타내고 벡터 연산으로 유사도를 계산하므로 수학적으로 대수학, 특히 선형대수에 근거한다.

오답 포인트

②의 부분 매칭이 가능하다는 사실을 ④의 의미 관계까지 잘 표현한다는 뜻으로 확대하면 안 된다. 벡터의 유사도 계산은 가능하지만, 기본 모델 자체가 동의어나 어구의 의미를 이해하는 것은 아니다.

NoteBase

다시 나오면 이렇게 판단

‘용어 벡터·유사도 계산·부분 매칭·순위화’는 벡터공간 모델의 특징으로 판단한다. 반면 ‘동의어·어구 등 의미 관계를 직접 표현하기 쉽다’는 설명은 기본 벡터공간 모델의 한계와 반대이므로 틀린 설명으로 판단한다.

암기 포인트

벡터공간 모델은 벡터로 비교하고 순위화하지만, 의미 관계를 직접 이해하지는 못한다.

NoteBase

12. 대상 영역과 표준 메타데이터 스키마가 바르게 연결된 것은?

① 도서관 - VRA
② 저작권 - ONIX
③ 박물관 - EAD
④ 기록물 - GEM
⑤ 교육 - LOM

정답 5번

출제 포인트

표준 메타데이터 스키마가 어떤 대상 자원을 기술하기 위해 만들어졌는지를 구분하는 문제이다.

핵심 개념

VRA Core는 시각문화 작품과 그 이미지를 기술하는 메타데이터 스키마이다.
ONIX는 출판사가 서적 등 출판 상품의 정보를 유통하기 위한 표준이다.
EAD는 기록물 자체가 아니라 기록물 검색도구를 구조화하여 부호화하는 표준이다.
GEM은 교육 자료의 검색과 기술을 위한 메타데이터 요소 집합이다.
LOM은 학습 객체의 내용, 교육적 특성, 기술적 특성 등을 기술하는 표준 메타데이터 스키마이다.

선지별 해설

① X — 틀린 표현은 '도서관 - VRA'이다. VRA Core의 대상은 일반적인 도서관 자료가 아니라 시각문화 작품과 그 이미지이므로 '시각자료 - VRA Core'로 고쳐야 한다.
② X — 틀린 표현은 '저작권 - ONIX'이다. ONIX는 저작권 전용 스키마가 아니라 출판 상품 정보를 유통하기 위한 표준이므로 '출판 유통 정보 - ONIX'로 고쳐야 한다.
③ X — 틀린 표현은 '박물관 - EAD'이다. EAD는 박물관 소장품 일반이 아니라 기록물의 검색도구를 부호화하므로 '기록물 검색도구 - EAD'로 고쳐야 한다.
④ X — 틀린 표현은 '기록물 - GEM'이다. GEM은 기록물용 스키마가 아니라 교육 자료를 기술하므로 '교육 자료 - GEM'으로 고쳐야 한다.
⑤ O — LOM은 Learning Object Metadata의 약자로, 학습 객체와 교육 자원을 기술하는 표준이므로 '교육 - LOM'의 연결이 옳다.

오답 포인트

GEM과 LOM이 모두 교육 자원과 관련된다는 점, EAD가 기록물 자체보다 기록물 검색도구를 기술한다는 점을 놓치기 쉽다. 스키마의 명칭만 보고 기관 유형과 연결하면 ①이나 ③을 고르기 쉽다.

NoteBase

다시 나오면 이렇게 판단

약어를 대상 자원과 함께 연결한다. VRA Core는 시각자료, ONIX는 출판 상품, EAD는 기록물 검색도구, GEM은 교육 자료, LOM은 학습 객체로 판단한다.

암기 포인트

VRA-시각자료, ONIX-출판, EAD-기록물 검색도구, GEM-교육 자료, LOM-학습 객체이다.

NoteBase

13. 온톨로지에 대한 설명으로 <보기>에서 옳은 것만을 모두 고르면?

<보 기>

ㄱ. 개념들 간 다양한 의미관계를 표현하고, 이러한 의미를 토대로 추론이 가능하도록 구성된다.

ㄴ. 개념과 관계에 대한 추론을 위해 논리적인 규칙을 명시한다.

ㄷ. 니콜라 구아리노(N. Guarino)는 일반성 수준에 따라 온톨로지를 상위수준 온톨로지, 하위수준 온톨로지, 도메인 온톨로지의 3가지 유형으로 구분하였다.

ㄹ. 온톨로지 언어의 구성요소 중 하나인 클래스는 도메인 내의 실체를 나타내는 개념을 정의한다.

① ㄱ, ㄴ

② ㄱ, ㄹ

③ ㄱ, ㄴ, ㄹ

④ ㄴ, ㄷ, ㄹ

⑤ ㄱ, ㄴ, ㄷ, ㄹ

정답 3번

출제 포인트

온톨로지의 구성 원리, 추론 기능, 유형 분류, 클래스의 의미를 구분하여 옳은 설명의 조합을 찾는 문제이다.

핵심 개념

온톨로지는 특정 영역의 개념과 개념 사이의 관계를 명시적으로 표현한 지식 구조이다. 개념은 클래스로, 구체적인 대상은 인스턴스로 나타내며, 관계와 공리·규칙을 이용해 새로운 지식을 추론할 수 있도록 구성한다. 일반성 수준에 따른 구아리노의 분류는 상위수준 온톨로지, 도메인 온톨로지, 과업 온톨로지, 응용 온톨로지의 네 유형이다.

NoteBase

선지별 해설

① X — ㄱ과 ㄴ은 옳지만, 옳은 ㄹ이 빠져 있다. 올바른 조합은 ㄱ, ㄴ, ㄹ이다.
② X — ㄱ과 ㄹ은 옳지만, 옳은 ㄴ이 빠져 있다. 올바른 조합은 ㄱ, ㄴ, ㄹ이다.
③ O — ㄱ, ㄴ, ㄹ이 옳고 ㄷ이 틀리므로 옳은 것만 모두 고른 조합이다. ㄱ은 개념 사이의 다양한 의미관계를 표현하고 이를 바탕으로 추론할 수 있도록 온톨로지를 구성한다는 설명이다. ㄴ은 개념과 관계에 관한 공리·논리적 규칙을 명시하여 추론을 지원한다는 설명이다. ㄹ은 클래스가 도메인 안에서 공통 속성을 지닌 인스턴스들의 집합에 해당하는 개념을 정의한다는 설명이다.
④ X — ㄴ과 ㄹ은 옳지만 ㄷ이 포함되어 있고, 옳은 ㄱ이 빠져 있다. ㄷ의 '상위수준·하위수준·도메인의 3가지'가 틀렸으며, 올바르게는 상위수준·도메인·과업·응용 온톨로지의 네 유형으로 구분한다. 올바른 조합은 ㄱ, ㄴ, ㄹ이다.
⑤ X — ㄱ, ㄴ, ㄹ은 옳지만 틀린 ㄷ까지 포함되어 있다. ㄷ은 세 유형이 아니라 상위수준·도메인·과업·응용 온톨로지의 네 유형으로 고쳐야 한다. 올바른 조합은 ㄱ, ㄴ, ㄹ이다.

오답 포인트

ㄷ은 '하위수준 온톨로지'를 독립 유형으로 제시하고 전체를 세 유형이라고 한 부분이 함정이다. 구아리노의 분류에서는 도메인 온톨로지 외에 과업 온톨로지와 응용 온톨로지를 구별하므로 네 유형으로 정리해야 한다.

다시 나오면 이렇게 판단

온톨로지 문항에서는 먼저 '개념·관계의 명시적 표현과 추론'이라는 기능을 확인한 뒤, 구성요소는 클래스·인스턴스·관계·공리로 구분한다. 구아리노의 일반성 수준 분류가 나오면 상위수준·도메인·과업·응용의 네 항목이 모두 있는지 확인한다.

암기 포인트

온톨로지는 개념과 관계를 명시하고 추론하며, 구아리노의 분류는 '상위·도메인·과업·응용' 네 가지이다. 클래스는 인스턴스들의 공통 개념이다.

NoteBase

14. SOAP(Simple Object Access Protocol)에 대한 설명으로 옳지 않은 것은?

① 모든 SOAP 메시지는 필수적인 SOAP 패키지(envelope), 선택적인 SOAP 헤더(header), 그리고 선택적인 SOAP 바디(body)로 구성된다.
② 용어의 의미로 보면 단순객체접근 프로토콜이다.
③ 기능적으로 보면 XML과 HTTP 등을 기반으로 하여 다른 컴퓨터에 있는 데이터나 서비스를 호출하기 위한 통신규약이다.
④ SOAP 패키지는 메시지를 표현하는 XML 문서의 최상위 요소이다.
⑤ SOAP 헤더는 통신 주체들 간 사전 합의 없이, 분산화된 방법으로 특성들을 SOAP 메시지에 추가하기 위한 일반적인 메커니즘이다.

정답 1번

출제 포인트

SOAP의 기본 의미와 메시지 구조에서 Envelope·Header·Body의 필수 여부 및 역할을 구별하는 문제이다.

핵심 개념

SOAP는 XML 형식의 메시지를 이용하여 서로 다른 시스템 사이에서 데이터와 서비스를 교환하도록 하는 통신 프로토콜이다.
SOAP 메시지는 최상위 요소인 Envelope 안에 선택적인 Header와 필수적인 Body를 두는 구조이다.
Header는 부가 기능과 처리 정보를 담고, Body는 실제 전송할 핵심 내용을 담는다.

NoteBase

선지별 해설

① X
틀린 표현은 '선택적인 SOAP 바디(body)'이다. Body는 실제 메시지 내용을 담는 필수 요소이다. 올바르게는 '필수적인 SOAP Envelope, 선택적인 SOAP Header, 필수적인 SOAP Body로 구성된다'이다.

② O
SOAP는 제시된 명칭인 Simple Object Access Protocol을 우리말로 옮기면 단순 객체 접근 프로토콜이라는 의미이다.

③ O
SOAP는 XML로 메시지를 표현하고 HTTP 등의 전송 수단을 이용하여 이기종 컴퓨터의 데이터나 서비스를 호출·교환할 수 있게 하는 통신 규약이다.

④ O
Envelope는 SOAP 메시지 전체를 감싸며 해당 XML 문서가 SOAP 메시지임을 나타내는 최상위 요소이다.

⑤ O
Header는 인증, 트랜잭션, 라우팅 같은 부가 처리 정보를 헤더 블록으로 확장하여 넣는 일반적인 메커니즘을 제공하므로 통신 주체 사이의 사전 합의에 메시지 전체 구조를 고정하지 않고도 분산된 방식으로 기능을 추가할 수 있다.

오답 포인트

①은 Header와 Body를 모두 선택적이라고 묶어 놓은 것이 함정이다. Header는 없어도 되지만 Body는 실제 메시지 내용을 담으므로 반드시 존재해야 한다.

다시 나오면 이렇게 판단

SOAP 구성 요소의 필수 여부를 먼저 확인한다. Envelope와 Body는 필수이고 Header만 선택적이라는 기준으로 선지를 판단한다.

암기 포인트

SOAP 구조는 'Envelope 필수, Header 선택, Body 필수'이다.

NoteBase

15. 메타데이터 영역에서 응용 프로파일(application profile)에 대한 설명으로 옳지 않은 것은?

① 특정한 응용 영역을 위해 정의된 메타데이터의 요소, 정책, 지침의 집합을 말한다.
② 해당 응용 영역에 필요한 새로운 메타데이터 요소들을 많이 개발하여 사용하는 것을 권장한다.
③ 응용시스템에서 사용되는 메타데이터 스키마를 결정하기 위해서는 메타데이터가 웹에서 사용 가능해야 한다.
④ 하나 이상의 네임스페이스(namespace)에서 추출한 메타데이터 요소들로 구성된 메타데이터 스키마를 말한다.
⑤ 메타데이터 요소에 대해 필요 시 인증된 스킴(scheme)이나 값(values)을 가진다는 특성이 있다.

정답 2번

출제 포인트

응용 프로파일의 정의와 구성 원칙, 특히 기존 메타데이터 요소의 재사용 원칙을 묻는 문제이다.

핵심 개념

응용 프로파일(application profile)은 특정 응용 영역의 요구에 맞게 하나 이상의 메타데이터 네임스페이스에서 요소를 가져와 사용 규칙, 제약, 값 체계 등을 정리한 명세이다. 핵심은 상호운용성을 위해 검증된 기존 요소를 우선 재사용하고, 새 요소는 꼭 필요한 경우에만 정의한다는 점이다.
네임스페이스(namespace)는 메타데이터 요소의 이름 충돌을 막고 그 출처와 의미를 식별하게 하는 이름 공간이며, 스킴(scheme)은 요소에 사용할 값의 형식이나 통제어휘 등을 지정하는 체계이다.

NoteBase

선지별 해설

① O — 응용 프로파일은 특정 응용 영역을 위해 선택한 메타데이터 요소와 그 사용 정책·지침을 모아 놓은 명세이다.

② X — 틀린 표현은 '새로운 메타데이터 요소들을 많이 개발하여 사용하는 것을 권장한다'이다. 올바르게는 상호운용성을 높이기 위해 기존 네임스페이스의 검증된 요소를 최대한 재사용하고, 새 요소는 필요한 경우에 한정하여 정의한다.

③ O — 응용 프로파일의 메타데이터 스키마는 기계가 해석하고 공유할 수 있는 형태로 웹에서 접근 가능하게 제공하는 것이 원칙이다.

④ O — 응용 프로파일은 하나 이상의 네임스페이스에서 추출한 메타데이터 요소들을 조합해 구성한 메타데이터 스키마이다.

⑤ O — 응용 프로파일은 각 메타데이터 요소에 필요에 따라 허용되는 스킴이나 값, 의무성·반복성 같은 사용 제약을 지정할 수 있다.

오답 포인트

②의 함정은 응용 영역에 맞춘다는 말을 '새 요소를 많이 만든다'는 뜻으로 바꾼 데 있다. 응용 프로파일의 맞춤화는 기존 요소의 선택·조합과 사용 제약의 구체화가 중심이지, 독자 요소의 대량 개발이 중심이 아니다.

다시 나오면 이렇게 판단

응용 프로파일 관련 선지는 '특정 목적에 맞춘 명세', '여러 네임스페이스 요소의 재사용·조합', '사용 규칙과 값 제약의 명시'이면 옳은 방향이다. '새 요소를 많이 개발하도록 권장한다'처럼 상호운용성을 떨어뜨리는 표현이면 틀린 것으로 판단한다.

암기 포인트

응용 프로파일은 새로 많이 만드는 것이 아니라, 기존 요소를 골라 묶고 사용법을 좁혀 정하는 명세이다.

NoteBase

16. 다음 빈칸에 공통으로 들어갈 알맞은 용어는?

> ○ 연구 데이터가 국가와 기관의 자산이라는 인식이 확산되면서 체계적인 데이터 관리와 재사용을 보장하기 위해서 연구자들에게 ( )이/가 요구되기 시작했다.
>
> ○ ( )의 주요 내용은 연구 과정에서 생산할 데이터, 데이터 유형과 형식, 데이터에 대한 장기적 접근 방법, 데이터 재사용성 보장 방법, 데이터를 설명하는 메타데이터 표준, 데이터 저장과 백업 위치, 데이터 공유를 위한 데이터 출판 방법 등으로 구성되어 있다.

① Open Data
② Open Science
③ Data Publishing
④ Data Management Plan
⑤ Research Data Management

정답 4번

출제 포인트

연구데이터의 생산부터 보존·공유까지 관리 방식을 미리 정한 문서가 데이터 관리 계획(Data Management Plan)임을 구별하는 문제이다.

핵심 개념

Data Management Plan(DMP)은 연구 과정에서 만들거나 수집할 데이터의 유형과 형식, 접근 방법, 메타데이터, 저장·백업, 장기 보존, 재사용 및 공유 방법 등을 사전에 정리한 계획 문서이다.
Research Data Management(RDM)은 연구데이터의 수집·정리·저장·보존·공유 등 데이터 생애주기 전반을 관리하는 활동과 체계를 뜻한다. 즉 DMP는 계획 문서이고, RDM은 그 계획을 포함하여 실제로 수행하는 더 넓은 관리 활동이다.

NoteBase

선지별 해설

① X — Open Data는 누구나 데이터에 접근하여 이용·재이용·배포할 수 있도록 개방하는 것을 뜻한다. 빈칸의 장기 접근, 저장, 백업, 메타데이터, 공유 방식을 종합한 계획 문서를 가리키지는 않는다. 틀린 표현인 Open Data는 Data Management Plan으로 고쳐야 한다.
② X — Open Science는 연구 과정과 성과를 개방하고 공유하려는 포괄적 연구 방식이다. 개별 연구의 데이터 관리 방법을 사전에 명시하는 문서 자체가 아니다. 틀린 표현인 Open Science는 Data Management Plan으로 고쳐야 한다.
③ X — Data Publishing은 데이터를 설명 정보와 함께 공개하여 다른 사람이 발견하고 이용할 수 있게 하는 출판·공개 과정이다. 데이터의 생산부터 저장·백업·장기 보존까지 정하는 종합 계획은 아니다. 틀린 표현인 Data Publishing은 Data Management Plan으로 고쳐야 한다.
④ O — Data Management Plan은 연구에서 생산할 데이터의 유형·형식, 장기 접근, 재사용, 메타데이터, 저장·백업 위치 및 공유·출판 방법을 정리한 계획이다. 제시문의 두 빈칸에 공통으로 들어간다.
⑤ X — Research Data Management는 연구데이터의 생애주기 전반을 관리하는 활동과 체계라는 넓은 개념이다. 제시문처럼 구체적인 관리 항목으로 구성된 계획 문서는 Data Management Plan이다. 틀린 표현인 Research Data Management는 Data Management Plan으로 고쳐야 한다.

오답 포인트

⑤ Research Data Management도 데이터의 저장·보존·공유를 다루므로 가장 헷갈리기 쉽다. 그러나 제시문은 관리 활동 전반이 아니라 연구 전에 작성하여 관리 방법을 명시하는 '계획'의 구성 내용을 열거하고 있다.

다시 나오면 이렇게 판단

데이터 유형·형식, 메타데이터, 저장·백업, 보존, 접근, 재사용, 공유 방법을 사전에 문서화한다는 표현이 함께 나오면 Data Management Plan으로 판단한다. 실제 관리 활동 전반을 묻는다면 Research Data Management로 구별한다.

암기 포인트

DMP는 데이터 관리의 계획 문서, RDM은 데이터 생애주기 관리 활동이다.

NoteBase

17. 한 문헌의 색인어가 <보기>와 같을 때, 이 문헌을 검색하는 불리언 연산식으로 옳지 않은 것은?

<보 기>

국회도서관; 대출; 베스트셀러; 분석; 빅데이터; 수서; 연구; 이용; 파이썬;

① 국회도서관 AND 빅데이터
② 빅데이터 AND (대출 OR 반납)
③ 베스트셀러 AND 분석 NOT 공공도서관
④ (파이썬 OR 자바) AND (연구 NOT 빅데이터)
⑤ (국회도서관 AND 이용) OR (대출 NOT 베스트셀러)

정답 4번

출제 포인트

문헌의 색인어 포함 여부에 AND, OR, NOT을 적용하여 각 불리언 연산식이 해당 문헌을 검색하는지 판단하는 문제이다.

핵심 개념

AND는 연결된 검색어가 모두 문헌에 있어야 참이 되고, OR는 하나 이상만 있어도 참이 된다. NOT은 뒤의 검색어가 있는 문헌을 제외한다. 괄호가 있으면 괄호 안을 먼저 계산하며, 이 문헌에는 '국회도서관, 대출, 베스트셀러, 분석, 빅데이터, 소서, 연구, 이용, 파이썬'이 포함되어 있다.

선지별 해설

① O — '국회도서관'과 '빅데이터'가 모두 색인어에 있으므로 AND 조건이 참이다.
② O — '빅데이터'가 있고, 괄호 안에서는 '대출'이 있으므로 '대출 OR 반납'이 참이다. 따라서 전체 AND 조건도 참이다.
③ O — '베스트셀러'와 '분석'이 모두 있고, 제외어인 '공공도서관'은 색인어에 없다. 따라서 '베스트셀러 AND 분석' 결과에서 '공공도서관'을 제외해도 이 문헌은 남는다.
④ X — 틀린 부분은 '연구 NOT 빅데이터'이다. '연구'가 있더라도 이 문헌에는 제외어인 '빅데이터'도 있으므로 괄호 안이 거짓이 되고, 전체 연산식으로는 이 문헌을 검색할 수 없다. 올바르게는 '(파이썬 OR 자바) AND 연구'처럼 실제 포함된 '연구'를 요구하면서 '빅데이터'를 제외하지 않아야 한다.
⑤ O — 첫 번째 괄호의 '국회도서관 AND 이용'이 참이다. 두 번째 괄호는 '베스트셀러' 때문에 거짓이지만, 두 괄호가 OR로 연결되어 있으므로 전체 식은 참이다.

NoteBase

오답 포인트

④에서 '연구'와 '빅데이터'가 모두 색인어에 있다는 사실만 보고 참으로 판단하기 쉽다. 그러나 NOT은 뒤의 검색어까지 포함하라는 뜻이 아니라 그 검색어가 포함된 문헌을 제외한다는 뜻이므로, 'NOT 빅데이터'가 이 문헌을 탈락시킨다.

다시 나오면 이렇게 판단

각 색인어의 존재 여부를 참과 거짓으로 바꾼 뒤 괄호부터 계산한다. AND는 모두 참, OR는 하나 이상 참, 'A NOT B'는 A가 참이고 B가 거짓일 때만 참으로 판단한다.

암기 포인트

AND는 모두 포함, OR는 하나 이상 포함, NOT은 뒤의 검색어를 제외한다.

NoteBase

18. 더블린 코어(Dublin Core)의 15개 기본요소에 해당하지 않는 것은?

① Title(표제)
② Location(위치)
③ Description(설명)
④ Source(출처)
⑤ Coverage(범위)

정답 2번

출제 포인트

더블린 코어의 15개 기본요소에 포함되는 메타데이터 항목을 구별하는 문제이다. 문제의 '해당하지 않는'이라는 부정형 표현에 주의해야 한다.

핵심 개념

더블린 코어는 정보자원의 검색과 식별을 돕기 위해 표제, 주제, 설명, 출처 등 15개의 기본 메타데이터 요소를 제시한 표준이다. 15개 요소는 Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights이다. Location은 이 목록에 포함되지 않는다.

선지별 해설

① O — Title(표제)은 정보자원에 부여된 이름을 나타내는 기본요소이다.
② X — Location(위치)은 더블린 코어의 15개 기본요소가 아니다. 틀린 표현은 'Location'이며, 위치·공간적 범위와 관련된 기본요소로는 Coverage(범위)가 제시된다.
③ O — Description(설명)은 정보자원의 내용에 대한 설명이나 요약을 나타내는 기본요소이다.
④ O — Source(출처)는 현재 정보자원이 유래한 관련 자원을 나타내는 기본요소이다.
⑤ O — Coverage(범위)는 정보자원의 공간적·시간적 범위 등을 나타내는 기본요소이다.

오답 포인트

'위치'가 메타데이터로 자연스럽게 보이기 때문에 Location을 기본요소로 착각하기 쉽다. 그러나 더블린 코어의 공식 요소명은 Location이 아니라 Coverage이며, 문제는 포함되는 항목이 아닌 것을 묻는다.

다시 나오면 이렇게 판단

선지의 영문 명칭을 15개 기본요소의 고정 목록과 대조한다. 위치와 관련된 표현이 나오면 독립 요소인 Location이 아니라 공간적·시간적 범위를 나타내는 Coverage인지 확인한다.

NoteBase

암기 포인트

더블린 코어에는 Coverage는 있고 Location은 없다.

NoteBase

19. <보기>에 제시된 XML 구문에 대한 설명으로 옳지 않은 것은?

<보 기>

```
<mods version="3.5"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-5.xsd">
<titleInfo>
  <title>Sound and fury</title>
  <subTitle>the making of the punditocracy</subTitle>
</titleInfo>
<name type="personal"
authorityURI="http://id.loc.gov/authorities/names"
valueURI="http://id.loc.gov/authorities/names/n92101908">
   <namePart>Alterman, Eric</namePart>
   <role>
    <roleTerm type="text">creator</roleTerm>
   </role>
</name>
<typeOfResource>text</typeOfResource>
<genre authority="marcgt">bibliography</genre> ...(하략)...
```

① 미 의회도서관에 의해 공식 발표되고 관리·유지되며, XML 스키마 기반의 인코딩 구조를 채택하고 있다.

② 상위 요소와 그에 따른 하위 요소로 구성되어 계층 구조를 가지며 속성은 각각의 요소에 적용된다.

③ 이 XML이 기술하는 자원의 표제는 본표제와 기타 표제 관련 정보를 포함한다.

④ Alterman, Eric은 기술하는 자원의 저자로 KORMARC 형식으로는 110필드에 해당한다.

⑤ typeOfResource 요소는 더블린 코어(Dublin Core)의 Type 요소로 대응될 수 있다.

정답 4번

출제 포인트

이 문제는 MODS의 XML 구조와 주요 요소의 의미, 그리고 MODS 요소를 KORMARC 및 더블린 코어와 대응시키는 방법을 묻는다.

NoteBase

핵심 개념

MODS는 미 의회도서관이 관리하는 XML 스키마 기반의 서지 메타데이터 형식이다. XML에서는 상위 요소 안에 하위 요소가 계층적으로 배치되고, 속성은 시작 태그 안에서 해당 요소에 부가 정보를 준다. 보기의 titleInfo는 표제 정보를, name은 이름과 역할 정보를, typeOfResource는 자원의 유형을 나타낸다. KORMARC에서 개인명 기본표목은 100필드이고, 110필드는 단체명 기본표목이다.

선지별 해설

① O — MODS는 미 의회도서관에 의해 공식적으로 발표·관리되는 XML 스키마 기반의 메타데이터 형식이다. 보기에도 mods 루트 요소와 스키마 위치가 제시되어 있다.

② O — MODS 레코드는 mods 아래에 titleInfo, name, typeOfResource 같은 상위 요소와 title, subTitle, namePart, roleTerm 같은 하위 요소가 놓이는 계층 구조이다. version, type, authorityURI 같은 속성도 각각의 요소에 적용된다.

③ O — titleInfo 안의 title은 본표제를, subTitle은 부표제를 나타낸다. 따라서 이 자원의 표제 정보에는 본표제와 그 밖의 표제 관련 정보가 함께 포함된다.

④ X — 틀린 표현은 ‘KORMARC 형식으로는 110필드’이다. Alterman, Eric은 개인 저자이므로 KORMARC에서는 개인명 기본표목인 100필드에 해당한다. 110필드는 단체명 기본표목에 사용한다.

⑤ O — MODS의 typeOfResource는 자원의 일반적 유형을 나타내므로 더블린 코어의 Type 요소에 대응시킬 수 있다. 보기에서는 그 값이 text로 제시되어 있다.

오답 포인트

④의 함정은 KORMARC의 100필드와 110필드를 바꾸어 제시한 데 있다. 사람 이름이라는 사실만 확인하면 개인명 100필드와 단체명 110필드를 구별할 수 있다.

다시 나오면 이렇게 판단

먼저 XML의 요소명과 값으로 기술 대상을 확인한 뒤, 다른 메타데이터 형식의 대응 필드를 판단한다. KORMARC 기본표목은 개인명 100, 단체명 110이라는 구분을 적용한다.

암기 포인트

KORMARC 기본표목은 개인명 100, 단체명 110이다. MODS의 typeOfResource는 더블린 코어의 Type에 대응한다.

NoteBase

20. 문헌요약의 유형으로서 초록에 대한 설명으로 옳지 않은 것은?

① 문헌의 생산자가 직접 작성하는 초록을 저자초록이라고 한다.
② 초록은 최신정보를 주지시키는 역할과 함께 독서 시간을 단축시키고 문헌검색을 용이하게 한다.
③ 초록은 별도의 초록지 또는 서지의 한 형태로 작성될 수 있다.
④ 통보적(informative) 초록은 원문헌의 대용물로 취급될 수 있다.
⑤ 지시적(indicative) 초록은 정보자료의 목적과 범위, 사용된 방법, 결과, 결론을 필수요소로 포함해야 한다.

정답 5번

출제 포인트

이 문제는 초록의 일반적 기능과 저자초록·통보적 초록·지시적 초록의 특징을 구분하는 문제이다.

핵심 개념

초록은 문헌의 핵심 내용을 간결하게 제시하여 최신정보의 파악, 독서 시간의 단축, 문헌검색과 관련성 판단을 돕는 요약문이다.
저자초록은 원문헌의 저자가 직접 작성한 초록이다.
통보적 초록은 목적·방법·결과·결론 등 원문헌의 핵심 정보를 구체적으로 전달하므로 경우에 따라 원문헌의 대용물로 활용될 수 있다.
지시적 초록은 문헌이 다루는 주제, 목적, 범위와 내용의 성격을 알려 원문헌의 필요성을 판단하게 하지만 결과와 결론까지 상세히 요약하지는 않는다.

선지별 해설

① O — 문헌의 생산자인 저자가 직접 작성한 초록을 저자초록이라고 한다.
② O — 초록은 문헌의 핵심을 압축하여 최신정보를 신속히 파악하게 하고, 원문을 모두 읽는 시간을 줄이며, 검색 단계에서 문헌의 관련성을 판단하기 쉽게 한다.
③ O — 초록은 원문헌에 붙는 형태뿐 아니라 여러 문헌의 초록을 모은 별도의 초록지나 초록 간행물의 형태로도 작성·제공될 수 있다.
④ O — 통보적 초록은 목적, 방법, 주요 결과와 결론 등 원문헌의 핵심 내용을 비교적 완전하게 제시하므로 원문헌의 대용물로 취급될 수 있다.
⑤ X — 틀린 표현은 '지시적 초록'과 '목적·범위·방법·결과·결론을 필수요소로 포함'한다는 부분이다.
올바르게는 통보적 초록이 목적·방법·결과·결론 등의 핵심 정보를 제시하며, 지시적 초록은 주제·목적·범위와 내용의 성격을 알려 주되 결과와 결론을 상세히 제시하지 않는다.

NoteBase

오답 포인트

⑤는 지시적 초록도 목적과 범위를 알려 준다는 점을 이용해 통보적 초록의 상세 요소인 방법·결과·결론까지 한데 묶은 선지이다. '결과와 결론까지 구체적으로 전달하는가'를 기준으로 두 유형을 구분해야 한다.

다시 나오면 이렇게 판단

초록 유형을 판단할 때 작성 주체가 저자이면 저자초록으로 보고, 내용 면에서는 원문의 핵심 결과와 결론까지 전달하면 통보적 초록, 문헌이 무엇을 다루는지만 안내하면 지시적 초록으로 구분한다.

암기 포인트

통보적 초록은 핵심 내용과 결과를 '통보'하고, 지시적 초록은 원문에 무엇이 있는지를 '지시'한다.

NoteBase