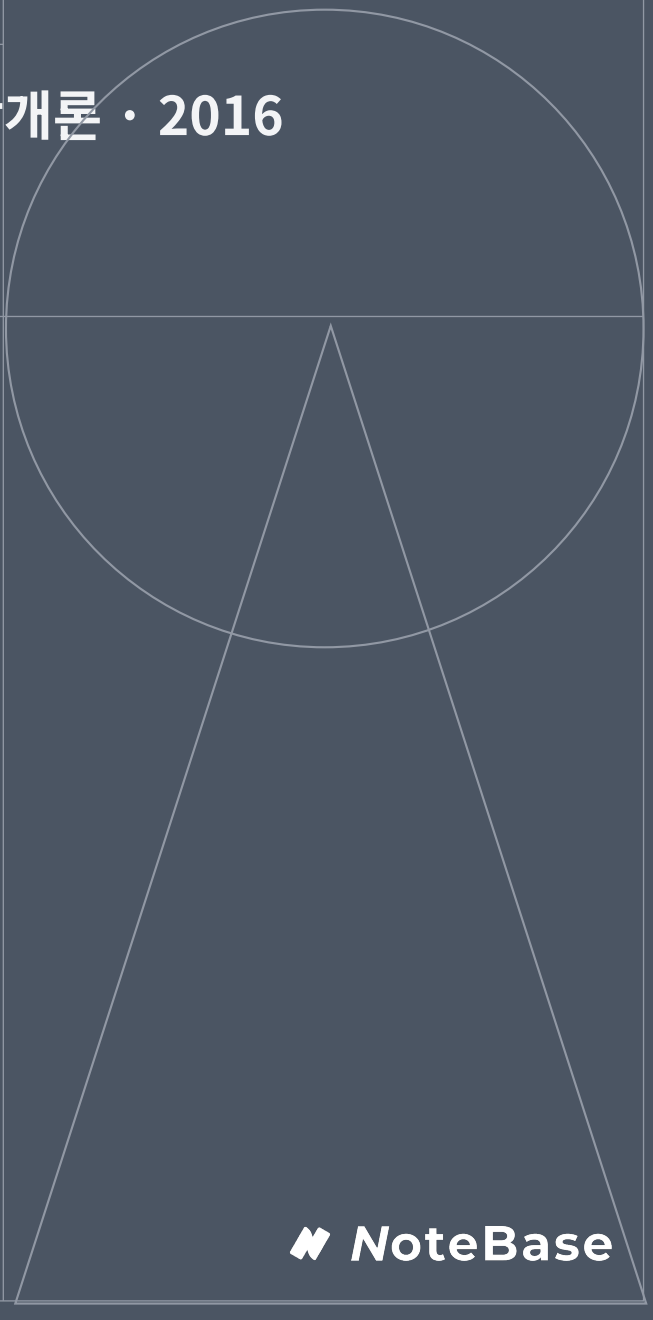
기출OX 해설집

공무원/국회직/9급 · 정보학개론 · 2016



NoteBase

기출OX 해설집

공무원/국회직/9급 · 정보학개론 · 2016

공무원/국회직/9급 · 정보학개론 · 2016

1. IFLA의 FRBR 모형에서 제시한 이용자 과업(User tasks)으로 옳지 않은 것은?

① 발견(find)
② 식별(identify)
③ 선정(select)
④ 입수(obtain)
⑤ 확인(clarify)

정답 5번

출제 포인트

이 문제는 FRBR 모형이 제시한 네 가지 이용자 과업을 구별하는 문제이다.

핵심 개념

FRBR의 이용자 과업은 발견(find), 식별(identify), 선정(select), 입수(obtain)이다. 발견은 조건에 맞는 자원을 찾는 일, 식별은 찾은 자원이 원하는 대상인지 확인하는 일, 선정은 이용 목적에 적합한 자원을 고르는 일, 입수는 선택한 자원에 접근하거나 이를 획득하는 일이다.

선지별 해설

① O — 발견(find)은 검색 조건에 맞는 개체나 자원을 찾아내는 FRBR 이용자 과업이다.
② O — 식별(identify)은 발견한 개체가 이용자가 찾던 대상인지 확인하고 유사한 개체와 구별하는 과업이다.
③ O — 선정(select)은 내용, 형식, 이용 가능성 등을 고려하여 이용 목적에 적합한 자원을 선택하는 과업이다.
④ O — 입수(obtain)는 선택한 자원에 접근하거나 대출·구입 등의 방법으로 이를 획득하는 과업이다.
⑤ X — 확인(clarify)은 FRBR이 제시한 이용자 과업이 아니다. 틀린 표현인 '확인(clarify)'은 FRBR의 네 과업인 발견(find), 식별(identify), 선정(select), 입수(obtain) 가운데 하나로 바뀌어야 한다.

오답 포인트

'확인'이라는 일상어가 식별과 비슷해 보여 FRBR의 정식 과업처럼 느껴지는 점이 함정이다. FRBR에서는 확인에 가까운 기능을 별도의 clarify로 두지 않고 식별(identify)로 표현한다.

NoteBase

다시 나오면 이렇게 판단

FRBR 이용자 과업을 묻는 문제에서는 '찾기→맞는 대상인지 확인하기→적합한 것 고르기→접근·획득하기'의 흐름에 각각 find, identify, select, obtain을 대응시킨다. 이 네 용어 밖의 표현은 FRBR의 정식 이용자 과업인지 의심한다.

암기 포인트

FRBR 이용자 과업은 발견·식별·선정·입수, 즉 FISO이다.

NoteBase

2. **<보기>는 검색시스템이 단어 "국회도서관"을 자동색인 한 결과이다. 이러한 자동색인 방법에 대한 설명 중 옳은 것은? ('_'는 문자열 전후에 오는 공란임)**

<보 기>

_국, 국회, 회도, 도서, 서관, 관_

① 다양한 구문형식의 단어구를 추출하여 최종 색인어 선정에 도움을 준다.
② 문헌에 사용된 단어의 의미적인 측면을 분석할 수 있어 검색 성능 향상이 가능하다.
③ 키워드로 자주 사용되는 미등록어를 등록하면 검색이 더 효율적이다.
④ 어절의 빈도를 이용하여 색인어를 추출하는 방법이다.
⑤ 문헌에 출현한 단어들의 철자 오류 발견 및 수정이 용이하다.

정답 5번

출제 포인트

문자열을 일정한 길이의 연속된 문자 조각으로 나누어 색인하는 n-gram 자동색인의 원리와 특징을 묻는 문제이다.

핵심 개념

n-gram 색인은 문자열에서 연속된 n개의 문자를 한 칸씩 이동하며 추출해 색인어로 삼는 방법이다. 보기의 '_국, 국회, 회도, 도서, 서관, 관_'은 단어의 앞뒤 경계를 '_'로 표시하고 두 문자씩 겹쳐 나눈 2-gram 결과이다. 사전의 단어 등록이나 문장의 구문·의미 분석에 의존하지 않으므로 미등록어와 철자 변형에도 비교적 유연하다.

NoteBase

선지별 해설

① X — 틀린 부분은 '다양한 구문형식의 단어구를 추출'한다는 표현이다. 이 방법은 구문 구조를 분석해 단어구를 고르는 것이 아니라 인접한 고정 길이의 문자 조각을 기계적으로 추출한다.
② X — 틀린 부분은 '단어의 의미적인 측면을 분석'한다는 표현이다. n-gram 색인은 문자열의 표면적인 문자 배열을 이용하며, 단어의 의미를 분석하는 방식은 아니다. 올바르게는 문자 단위의 부분 일치를 이용해 검색한다.
③ X — 틀린 부분은 '자주 사용되는 미등록어를 등록하면'이라는 표현이다. n-gram 색인은 사전에 등록되지 않은 말도 고정 길이의 문자열로 분할할 수 있으므로 미등록어의 키워드 등록을 전제로 하지 않는다. 올바르게는 검색에서 제외할 불용 문자열을 관리하면 불필요한 색인량을 줄이는 데 도움이 될 수 있다.
④ X — 틀린 부분은 '어절의 빈도를 이용'한다는 표현이다. 보기에서는 어절 빈도를 계산한 것이 아니라 '국회도서관'을 앞뒤 경계 표시를 포함해 두 문자씩 겹쳐 분할했다. 올바르게는 고정된 문자 길이와 위치 이동을 이용해 색인어를 추출한다.
⑤ O — 단어를 여러 개의 겹치는 문자 조각으로 저장하므로 철자가 일부 틀려도 나머지 조각의 일치 여부를 비교할 수 있다. 따라서 철자 오류의 발견과 수정 후보 탐색이 용이하다.

오답 포인트

보기의 '_'를 단순한 어절 공란으로 보고 ④를 고르기 쉽다. 그러나 '_'는 문자열의 시작과 끝을 나타내는 경계 표시이며, 핵심은 국회·회도·도서·서관처럼 두 문자씩 겹쳐 분할했다는 점이다.

다시 나오면 이렇게 판단

한 문자열이 'AB, BC, CD'처럼 일정한 길이로 겹쳐 분할되어 있으면 n-gram 색인으로 판단한다. 이 방식은 문자 배열을 처리하므로 구문 분석·의미 분석·어절 빈도 계산을 특징으로 제시한 선지는 배제한다.

암기 포인트

n-gram은 'n글자씩 겹쳐 자르기'이며, 미등록어와 철자 오류에 강하다.

NoteBase

3. <보기>에서 설명하는 내용을 제시한 정보학자는?

<보 기>

.논문의 인용 영향력 분석에 목적이 있다.

.과학기술분야에서 논문의 인용은 소수의 핵심 학술지에 수록된 논문에 집중된다는 개념이다.

.SCI(Science Citation Index) 학술지 중 1969년에 출판된 2,200종에 실린 논문 약 1백만 편에 대한 인용 분석을 실시하여 피인용논문의 24%가 25개 학술지에 집중적으로 게재되었다는 사실을 발견하였다.

① 조지 지프(George K. Zipf)
② 알프레드 로트카(Alfred J. Lotka)
③ 유진 가필드(Eugene Garfield)
④ 사무엘 브래드포드(Samuel C. Bradford)
⑤ 한스 룬(Hans P. Luhn)

정답 3번

출제 포인트

논문 인용이 소수의 핵심 학술지에 집중된다는 '집중의 법칙'과 이를 제시한 유진 가필드를 연결하는 문제이다.

핵심 정리

조지 지프: 단어의 사용 빈도는 빈도 순위에 반비례한다는 지프의 법칙을 제시하였다.
알프레드 로트카: 일정 수의 논문을 발표한 저자의 수에 관한 과학자 생산성의 법칙을 제시하였다.
유진 가필드: 인용색인 SCI를 창안하고, 여러 분야의 중요 논문이 소수의 핵심 학술지군에 집중된다는 집중의 법칙을 제시하였다.
사무엘 브래드포드: 특정 주제의 논문이 핵심 학술지와 여러 구역에 분산되는 양상을 설명하는 문헌분산의 법칙을 제시하였다.
한스 룬: KWIC 색인과 선택적 정보배포 등 정보검색의 자동화에 기여하였다.

NoteBase

선지별 해설

① X — 조지 지프와 연결한 부분이 틀렸다. 올바르게는 지프는 단어의 빈도와 빈도 순위의 관계를 설명한 지프의 법칙과 연결된다.
② X — 알프레드 로트카와 연결한 부분이 틀렸다. 올바르게는 로트카는 논문 수에 따른 저자 수의 분포를 설명한 과학자 생산성의 법칙과 연결된다.
③ O — 유진 가필드는 SCI를 창안했으며, SCI 인용 분석을 바탕으로 중요 논문이 소수의 핵심 학술지군에 집중된다는 집중의 법칙을 제시하였다. 보기에 제시된 1969년 SCI 학술지 및 인용 분석과 일치한다.
④ X — 사무엘 브래드포드와 연결한 부분이 틀렸다. 올바르게는 브래드포드는 특정 주제의 논문이 핵심 학술지와 그 밖의 구역에 분산되는 문헌분산의 법칙과 연결된다. 보기의 설명은 브래드포드의 분산 법칙을 확장하여 여러 분야에 공통되는 핵심 학술지군을 강조한 가필드와 연결된다.
⑤ X — 한스 룬과 연결한 부분이 틀렸다. 올바르게는 룬은 KWIC 색인과 선택적 정보배포 등 정보검색 자동화 기법과 연결된다.

출제자가 노린 함정

보기의 '소수의 핵심 학술지에 집중'이라는 표현만 보고 브래드포드의 문헌분산 법칙을 고르기 쉽다. 그러나 SCI를 이용한 인용 분석, 여러 과학기술 분야를 포괄하는 핵심 학술지군, 1969년 조사라는 단서의 결합은 가필드의 집중의 법칙을 가리킨다.

다시 나오면 이렇게 판단

SCI·인용색인·여러 분야에 공통되는 핵심 학술지군은 가필드로 판단한다. 특정 주제 문헌의 핵심 학술지와 구역별 분산은 브래드포드, 저자별 논문 생산량은 로트카, 단어 빈도와 순위는 지프로 구분한다.

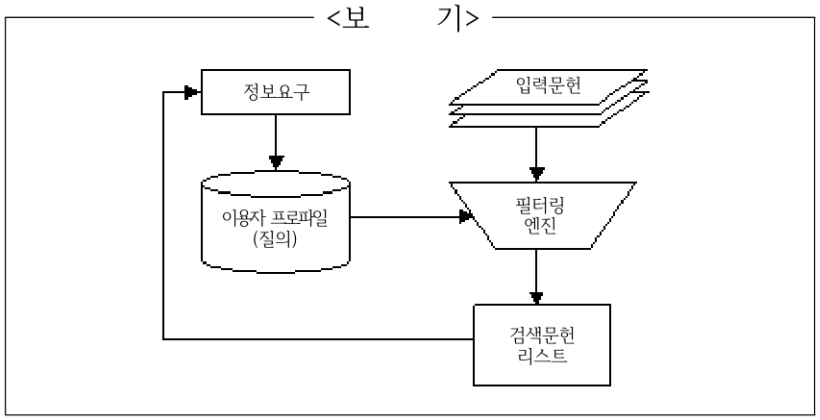
암기 포인트

가필드 = SCI·인용색인·집중, 브래드포드 = 특정 주제 문헌의 분산이다.

NoteBase

4. <보기>는 검색시스템의 필터링 과정을 보여주는 그림이다. 이에 대한 설명 중 옳지 않은 것은?

<보 기>



① 전자저널이나 학술 데이터베이스가 제공하는 'Alert' 서비스에 해당한다.

② 전통적인 도서관에서의 SDI(Selective Dissemination of Information) 서비스에 상응한다.

③ kNN 분류기, SVM 분류기, 신경망 분류기 등을 사용하여 검색여부를 결정할 수 있다.

④ 비교적 정적인 문헌집단을 대상으로 새로운 질의를 입력하는 검색환경에 적합하다.

⑤ 정보가 빠르고 다양하게 변화하는 최근의 웹 환경에 적용할 수 있다.

정답 4번

출제 포인트

이 문제는 도식에서 이용자의 정보요구로 만든 프로파일과 계속 들어오는 입력문헌을 필터링 엔진이 대조하고, 검색문헌 리스트에 대한 반응을 다시 프로파일에 반영하는 구조를 파악하는 문제이다. 정보 필터링은 지속적으로 유입되는 새 문헌 가운데 이용자의 관심에 맞는 것을 선별하는 방식이다.

NoteBase

선지별 해설

① O — Alert 서비스는 미리 설정한 주제나 검색 조건에 맞는 새 정보가 들어오면 이용자에게 알려 주므로 정보 필터링의 대표적 활용에 해당한다.

② O — SDI는 이용자의 관심 프로파일을 바탕으로 새로 입수된 문헌을 선택적으로 제공하는 서비스이므로 도식의 구조와 상응한다.

③ O — 필터링 엔진은 문헌이 이용자의 관심 범주에 속하는지를 판별해야 하므로 kNN, SVM, 신경망과 같은 분류 기법을 이용하여 검색 여부 또는 제공 여부를 결정할 수 있다.

④ X — '비교적 정적인 문헌집단을 대상으로 새로운 질의를 입력하는 검색환경'은 정보 필터링이 아니라 전통적인 정보검색에 적합한 설명이다. 올바르게는 정보 필터링은 이용자의 비교적 지속적인 관심을 프로파일로 유지하면서 새로 유입되는 동적인 문헌을 선별하는 환경에 적합하다.

⑤ O — 웹 환경은 정보가 빠르게 추가·갱신되고 형태도 다양하므로, 이용자 프로파일에 따라 새 정보를 계속 골라내는 정보 필터링을 적용하기에 적합하다.

오답 포인트

④의 '정적인 문헌집단'과 '새로운 질의'는 정보 필터링의 특징처럼 보이지만 실제로는 전통적 정보검색의 구조이다. 필터링은 질의를 매번 새로 만드는 방식보다 이용자 프로파일을 지속적으로 유지하고 새로 들어오는 문헌을 반복적으로 판별하는 데 초점이 있다.

다시 나오면 이렇게 판단

고정된 문헌집단에 새 질의를 적용하면 정보검색, 지속적인 이용자 프로파일에 새 문헌을 적용하면 정보 필터링으로 판단한다. 도식에 검색 결과의 피드백이 프로파일로 되돌아가는 화살표가 있으면 이용자의 관심을 갱신하는 적응형 필터링 구조로 본다.

암기 포인트

정보검색은 '새 질의-기존 문헌', 정보 필터링은 '기존 프로파일-새 문헌'의 대응이다.

NoteBase

5. 정보행태에 관한 모형의 설명으로 옳지 않은 것은?

① 테일러(R. Taylor)는 질의어 구성의 네 가지 단계를 설명한 '질문형성(Question formation)'단계 모형을 제시하였다.

② 벨킨(N. Belkin)은 지식의 생산 및 전달과정을 설명한 'ASK(Anomalous Status of Knowledge)'모형을 제시하였다.

③ 더빈(B. Dervin)은 문제해결 과정을 상황, 격차, 다리연결/이용의 세 가지 단계로 분석한 '의미형성(Sense-making)'모형을 제시하였다.

④ 윌슨(T. Wilson)은 '정보요구의 상황'에서 '정보처리와 이용'에 이르는 과정으로 구성된 정보행태모형을 제시하였다.

⑤ 레틱(J. Rettig)은 소음(noise)의 개념을 이용한 참고과정모델을 제시하였다.

정답 2번

출제 포인트

정보행태 연구의 주요 학자와 대표 모형·핵심 개념을 정확히 연결하는 문제이다.

핵심 정리

테일러(R. Taylor): 질문형성 단계 모형 — 이용자의 정보요구가 본능적 요구, 의식적 요구, 형식화된 요구, 타협된 요구의 네 수준으로 구체화되는 과정을 설명한다.
벨킨(N. Belkin): ASK 모형 — 이용자가 자신의 지식 상태에 이상이나 결핍이 있다고 인식할 때 정보요구와 정보탐색이 발생한다고 본다.
더빈(B. Dervin): 의미형성 모형 — 개인이 처한 상황에서 지식의 격차를 인식하고 이를 연결하여 정보를 이용하는 과정을 설명한다.
윌슨(T. Wilson): 정보행태모형 — 정보요구가 생기는 맥락에서 출발하여 정보탐색을 거쳐 정보처리와 이용으로 이어지는 과정을 설명한다.
레틱(J. Rettig): 참고과정모델 — 이용자와 참고사서 사이의 의사소통·피드백 과정에 소음이 개입할 수 있음을 반영한다.

NoteBase

선지별 해설

① O — 테일러는 정보요구가 질문으로 표현되는 과정을 네 수준의 질문형성 단계로 설명하였다.
② X — '지식의 생산 및 전달과정'이라는 설명이 잘못되었다. ASK는 이용자의 지식 상태에 이상이나 결핍이 있어 문제를 적절한 질문으로 표현하기 어려운 상태와 그로 인한 정보탐색을 설명하는 모형이다.
③ O — 더빈의 의미형성 모형은 상황에서 격차를 인식하고 다리를 놓듯 이를 연결하여 정보를 이용하는 과정을 설명한다.
④ O — 윌슨의 정보행태모형은 정보요구가 발생하는 맥락에서 정보탐색을 거쳐 정보처리와 이용으로 이어지는 흐름을 포함한다.
⑤ O — 레틱의 참고과정모델은 참고봉사를 의사소통 과정으로 파악하며, 전달을 방해하는 소음의 개념과 상호 피드백을 반영한다.

출제자가 노린 함정

②는 ASK의 영문 명칭을 정확히 제시해 맞는 선지처럼 보이게 한 뒤, 설명 부분을 '지식의 생산 및 전달과정'으로 바꾸었다. ASK의 핵심은 지식의 생산이 아니라 이용자가 느끼는 지식 상태의 이상·결핍이다.

다시 나오면 이렇게 판단

학자명과 모형명만 보지 않고 대표 문제의식을 함께 연결한다. 테일러는 정보요구의 질문화, 벨킨은 지식 상태의 이상, 더빈은 상황－격차－연결, 윌슨은 요구에서 처리·이용까지의 정보행태, 레틱은 참고과정의 의사소통과 소음으로 판단한다.

암기 포인트

테일러＝질문형성, 벨킨＝ASK·지식 결핍, 더빈＝상황－격차－연결, 윌슨＝요구－탐색－처리·이용, 레틱＝참고과정·소음이다.

NoteBase

6. 데이터 결합(Data fusion)이란 동일한 컬렉션에 대해 여러 다른 검색 관련 기법들을 적용하여 검색결과를 통합하는 전략이다. 다음 중 결합 대상이 되는 검색 관련 기법으로 옳지 않은 것은?

① 문헌 표현 방법의 결합
② 질의 형식의 결합
③ 용어 활용도의 결합
④ 검색 알고리즘의 결합
⑤ 검색 순위의 결합

정답 3번

출제 포인트

이 문제는 동일한 컬렉션에서 여러 검색 기법으로 얻은 결과를 통합하는 데이터 결합의 대상과, 단순한 용어 통계적 특성을 구별하는 문제이다.

핵심 개념

데이터 결합은 동일한 문헌 컬렉션에 서로 다른 문헌 표현, 질의 표현, 검색 알고리즘 등을 적용하여 얻은 검색 결과나 순위를 하나로 합치는 전략이다.
검색 기법의 결합: 검색 결과를 만들어 내는 표현 방식·질의 방식·검색 알고리즘 또는 그 결과 순위를 결합한다.
용어 활용도: 용어가 문헌이나 질의에서 사용되는 정도를 나타내는 특성으로, 그 자체는 독립된 검색 기법이 아니다.

선지별 해설

① O — 문헌 표현 방법을 달리하면 같은 문헌도 서로 다른 색인 특성으로 표현되어 검색 결과가 달라질 수 있으므로 결합 대상이 된다.
② O — 질의를 여러 형식으로 표현하거나 변형하여 얻은 검색 결과는 상호 보완적으로 결합할 수 있으므로 결합 대상이 된다.
③ X — 틀린 표현은 '용어 활용도'이다. 용어 활용도는 용어가 사용되는 정도를 나타내는 특성이지, 별도의 검색 결과를 산출하는 검색 기법이 아니다. 올바르게는 문헌 표현 방법, 질의 형식, 검색 알고리즘처럼 검색 결과를 산출하는 기법이나 그 결과가 결합 대상이 된다.
④ O — 서로 다른 검색 알고리즘은 각기 다른 기준으로 관련성을 계산하므로 그 결과를 결합할 수 있다.
⑤ O — 각 검색 기법이 산출한 검색 순위를 조정하거나 통합하는 것은 데이터 결합의 대표적인 방식이다.

오답 포인트

③의 '용어 활용도'를 용어 가중치나 질의 표현 기법과 혼동하기 쉽다. 이 문항에서는 별도의 검색 결과를 만들어 내는 기법인지가 판단 기준이며, 용어 활용도는 그러한 기법 자체가 아니다.

NoteBase

다시 나오면 이렇게 판단

선지가 문헌·질의의 표현 방식, 검색 알고리즘, 검색 결과의 순위처럼 독립적인 검색 결과를 만들거나 그 결과를 합치는 요소인지 확인한다. 단순한 용어의 사용 정도나 통계적 속성만 나타내면 검색 기법의 결합 대상으로 보지 않는다.

암기 포인트

데이터 결합은 표현·질의·알고리즘·순위를 합치며, 단순한 용어 활용도는 합치지 않는다.

NoteBase

7. 온톨로지(Ontology)에 대한 설명으로 옳은 것을 <보기>에서 모두 고른 것은?

<보 기>

ㄱ. 1980년대에 인공지능 연구자들이 전문가시스템 개발 중 한 영역을 개념화하면서 지식 습득과 표현에 대해 온톨로지라는 용어를 채택하였다.

ㄴ. 지식엔지니어링 또는 보다 광범하게 컴퓨터와 정보과학 분야에서, 온톨로지는 공유된 개념화에 대한 명시적이고 정형화된 명세서로 정의된다.

ㄷ. 온톨로지는 전통적인 통제어휘(예: 시소러스)와 같이 고정적이기 때문에 개념과 관계에 대한 추론을 위한 논리적 규칙을 명시해야만 한다.

ㄹ. 온톨로지에서 개념 사이의 관계는 동의어, 반의어, 하위어 그리고 부분어를 포함한다.

① ㄱ, ㄹ ② ㄴ, ㄹ

③ ㄱ, ㄴ, ㄷ ④ ㄱ, ㄴ, ㄹ

⑤ ㄱ, ㄴ, ㄷ, ㄹ

정답 4번

출제 포인트

온톨로지의 등장 배경과 정의, 전통적 통제어휘와의 차이, 온톨로지가 표현하는 개념 관계를 구별하는 문제이다.

핵심 개념

온톨로지는 특정 분야의 개념과 개념 사이의 관계를 컴퓨터가 처리할 수 있도록 명시적이고 형식적으로 표현한 지식 모델이다. 지식공학에서는 보통 '공유된 개념화에 대한 명시적이고 정형화된 명세'로 이해한다.
전통적 통제어휘나 시소러스는 용어를 통일하고 동의어·상하위어 등의 관계를 정리하는 데 초점이 있다.
온톨로지는 여기서 더 나아가 개념의 속성, 제약, 다양한 관계와 공리 등을 형식화하여 의미 기반 처리와 추론을 지원할 수 있다. 다만 모든 온톨로지가 별도의 추론용 논리 규칙을 반드시 명시해야 하는 것은 아니다.

NoteBase

선지별 해설

① X — ㄱ과 ㄹ은 옳지만, 온톨로지의 대표적 정의를 설명한 ㄴ도 옳으므로 옳은 설명을 모두 고른 조합이 아니다. 틀린 부분은 ㄴ을 제외한 것이며, 올바른 조합은 ㄱ, ㄴ, ㄹ이다.
② X — ㄴ과 ㄹ은 옳지만, 1980년대 인공지능 연구에서 전문가시스템의 지식 습득·표현 문제를 개념화하며 온톨로지라는 용어를 채택하였다는 ㄱ도 옳다. 틀린 부분은 ㄱ을 제외한 것이며, 올바른 조합은 ㄱ, ㄴ, ㄹ이다.
③ X — ㄱ과 ㄴ은 옳지만 ㄷ은 옳지 않고 ㄹ은 옳다. ㄷ의 '고정적이기 때문에 논리적 규칙을 명시해야만 한다'가 틀린 표현이다. 올바르게는 온톨로지가 통제어휘보다 개념의 관계·제약·공리 등을 더 형식적으로 표현하여 추론을 지원할 수 있지만, 모든 온톨로지가 별도의 논리 규칙을 반드시 명시해야 하는 것은 아니다. 따라서 올바른 조합은 ㄱ, ㄴ, ㄹ이다.
④ O — ㄱ은 온톨로지 용어가 인공지능의 지식 표현 분야에 도입된 배경을, ㄴ은 온톨로지의 대표적 정의를, ㄹ은 온톨로지가 표현할 수 있는 개념 관계의 예를 각각 올바르게 설명한다.
⑤ X — ㄱ, ㄴ, ㄹ은 옳지만 ㄷ이 옳지 않으므로 모두 옳은 조합이 아니다. 틀린 부분은 ㄷ을 포함한 것이며, 올바른 조합은 ㄱ, ㄴ, ㄹ이다. ㄷ은 '온톨로지는 관계와 제약, 공리 등을 형식적으로 표현하여 추론을 지원할 수 있다'로 고쳐야 한다.

오답 포인트

ㄷ은 '온톨로지가 추론을 지원한다'는 일반적 특징에 '논리적 규칙을 명시해야만 한다'는 필수 조건을 섞어 놓았다. '반드시', '해야만 한다'와 같은 단정 표현이 정의보다 강한지 확인해야 한다. 또한 이 문항은 개별 진술의 O/X가 아니라 옳은 진술을 모두 포함한 조합을 묻기 때문에, 옳은 ㄱ이나 ㄴ을 빠뜨린 ①·②도 정답이 될 수 없다.

다시 나오면 이렇게 판단

온톨로지 문항에서는 먼저 '공유된 개념화의 명시적·형식적 명세'라는 정의를 잡고, 통제어휘·시소러스보다 관계와 제약을 풍부하게 형식화한다는 점을 확인한다. 그다음 추론을 '지원할 수 있다'는 설명과 모든 온톨로지가 별도 규칙을 '반드시 명시해야 한다'는 과도한 단정을 구별한다.

암기 포인트

온톨로지 = 공유된 개념화의 명시적·형식적 명세, 관계와 제약을 표현해 의미 처리와 추론을 지원하는 지식 모델이다.

NoteBase

8. 다음 중 정보검색 관련 용어에 대한 설명으로 옳지 않은 것은?

① 페이지 랭크(page rank) 기능이란 특정 단어의 출현 빈도에 따라서 검색결과를 순위화하는 기능을 말한다.

② 메타검색엔진(metasearch engine)이란 서로 다른 검색 엔진들을 연결시켜 다른 검색 엔진이 검색한 정보를 보여주는 검색 엔진을 말한다.

③ 디렉토리 브라우징(directory browsing)이란 웹상의 문서들을 주제별로 범주화하여 원하는 정보를 찾아가도록 하는 서비스를 말한다.

④ 절단(truncation) 기법이란 용어의 일부분을 생략하고 나머지 부분만을 검색어로 사용하는 기법을 말한다.

⑤ 정확률(precision ratio)이란 검색된 문헌 중 적합한 문헌의 비율을 말한다.

정답 1번

출제 포인트

정보검색에서 사용하는 PageRank, 메타검색엔진, 디렉터리 브라우징, 절단, 정확률의 정의를 구별하는 문제이다.

핵심 개념

PageRank: 웹페이지 사이의 링크 구조를 분석하여 페이지의 상대적 중요도를 평가하고 검색 결과의 순위 결정에 활용하는 방식이다. 키워드는 링크 구조와 중요도이다.
메타검색엔진: 이용자의 질의를 여러 검색엔진에 보내고 각 검색 결과를 모아 제시하는 검색엔진이다. 키워드는 복수 검색엔진의 통합이다.
디렉터리 브라우징: 웹 문서를 주제별·계층별 범주로 분류하여 원하는 범주를 따라가며 정보를 찾게 하는 서비스이다. 키워드는 주제별 분류이다.
절단: 검색어의 어근이나 일부 문자열에 절단 기호를 붙여 형태가 다른 여러 단어를 함께 검색하는 기법이다. 키워드는 어근과 변형어 검색이다.
정확률: 검색된 문헌 전체 중 실제로 적합한 문헌이 차지하는 비율이다. 키워드는 검색 결과의 적합성이다.

NoteBase

선지별 해설

① X — 틀린 표현은 '특정 단어의 출현 빈도에 따라서 검색결과를 순위화'한다는 부분이다. PageRank는 단어의 출현 빈도가 아니라 웹페이지 사이의 링크 구조와 링크를 제공하는 페이지의 중요도 등을 분석하여 페이지의 상대적 중요도를 평가한다.
② O — 메타검색엔진은 하나의 질의를 서로 다른 여러 검색엔진에 전달하고, 그 검색 결과를 모아 이용자에게 보여 준다.
③ O — 디렉터리 브라우징은 웹 문서를 주제별·계층별 범주로 조직하여 이용자가 범주를 따라 원하는 정보를 찾도록 하는 방식이다.
④ O — 절단은 검색어의 어근 등 일부만 남기고 절단 기호를 사용하여 동일한 어근을 가진 여러 변형어를 함께 검색하는 기법이다. 더 정확히는 단순히 임의의 부분을 생략하는 것이 아니라, 검색하려는 공통 문자열과 절단 기호를 이용한다.
⑤ O — 정확률은 검색된 문헌 중 적합한 문헌의 비율이다. 즉 '적합 문헌 수 ÷ 검색된 전체 문헌 수'로 판단한다.

오답 포인트

①의 '출현 빈도'는 단어 빈도에 기초한 관련성 판단과 PageRank를 혼동하게 만드는 표현이다. PageRank의 결정적 단서는 페이지 내용에 같은 단어가 몇 번 나오는지가 아니라 다른 페이지들과 맺는 링크 관계이다. 또한 정확률은 전체 적합 문헌 중 찾아낸 비율인 재현율과 구별해야 한다.

다시 나오면 이렇게 판단

용어마다 핵심 기준어를 대응시킨다. PageRank는 링크 구조, 메타검색엔진은 여러 검색엔진의 결과 통합, 디렉터리 브라우징은 주제별 분류, 절단은 어근과 절단 기호를 이용한 변형어 검색, 정확률은 검색된 결과 중 적합 문헌의 비율로 판단한다.

암기 포인트

PageRank는 빈도가 아니라 링크, 정확률은 검색된 것 중 적합한 것의 비율이다.

NoteBase

9. <보기>에서 설명하고 있는 용어는?

<보 기>

.1945년 Vannevar Bush가 “As We May Think”라는 제목의 사설에서 제안하였다.

.기계를 사용한 온라인 하이퍼텍스트 정보검색 시스템의 초기 원형이 된다.

.정보검색의 기술과 비전을 제시한 최초의 디지털 처리 기계라는 의미가 있다.

.컴퓨터 하드웨어, 소프트웨어의 진보, 언어의 개발, 분류와 색인 등의 정보 가공에 대한 기능적 발전을 예견하였다.

① ZATOCODING
② MEMEX
③ PARSER
④ SKOS
⑤ SPIDER

정답 2번

출제 포인트

이 문제는 버니바 부시(Vannevar Bush), 1945년, 「As We May Think」, 연상에 따른 정보 연결이라는 단서를 정보검색 개념과 연결하는 문제이다.

핵심 정리

MEMEX는 버니바 부시가 1945년 「As We May Think」에서 제안한 가상 정보 장치이다. 이용자가 자료를 연상적으로 연결하고 그 경로를 따라 다시 찾는다는 구상으로, 후대 하이퍼텍스트의 선구적 개념으로 평가된다. 문제의 결정적 연결은 ‘버니바 부시 → As We May Think → MEMEX’이다.

NoteBase

선지별 해설

① X — ZATOCODING은 특성을 코드화하고 코드를 중첩하여 자료를 검색하는 방식으로, 버니바 부시가 「As We May Think」에서 제안한 장치명이 아니다. 틀린 표현 'ZATOCODING'은 'MEMEX'로 고쳐야 한다.
② O — MEMEX는 버니바 부시가 1945년 「As We May Think」에서 제안한 가상 장치이며, 자료 사이의 연상적 연결이라는 발상 때문에 하이퍼텍스트 정보검색의 초기 구상으로 평가된다.
③ X — PARSER는 문장이나 프로그램 코드의 문법적 구조를 분석하는 구문 분석기이므로 보기의 역사적 정보검색 장치와 다르다. 틀린 표현 'PARSER'는 'MEMEX'로 고쳐야 한다.
④ X — SKOS는 분류표·시소러스 같은 지식조직체계를 웹에서 표현하기 위한 데이터 모델이므로 1945년 부시가 제안한 장치가 아니다. 틀린 표현 'SKOS'는 'MEMEX'로 고쳐야 한다.
⑤ X — SPIDER는 웹 링크를 따라 문서를 자동으로 수집하는 크롤러를 가리키므로 보기의 인물·연도·논문과 연결되지 않는다. 틀린 표현 'SPIDER'는 'MEMEX'로 고쳐야 한다.

출제자가 노린 함정

ZATOCODING, PARSER, SKOS, SPIDER도 정보의 처리·검색과 관련된 용어라서 혼동하기 쉽다. 그러나 '1945년', 'Vannevar Bush', 'As We May Think'가 함께 제시되면 MEMEX를 선택해야 한다.

다시 나오면 이렇게 판단

인물·저술·연도를 먼저 묶어 판단한다. '버니바 부시 → 1945년 As We May Think → 연상적 연결 → MEMEX'의 연결 고리가 확인되면 다른 정보처리 용어를 배제한다.

암기 포인트

Bush의 MEMEX는 연상적 연결을 통해 하이퍼텍스트를 예견한 구상이다.

NoteBase

10. Kupiec(1995) 등이 제시한 기계학습을 통한 요약 기법에서 문장 추출 기준으로서의 자질 유형으로 옳지 않은 것은?

① 문장길이
② 고정 단어구
③ 문단/문장 위치
④ 단어빈도
⑤ 주제어

정답 4번

출제 포인트

Kupiec(1995) 등의 학습 기반 추출 요약에서 문장을 요약문에 포함할지 판단하는 자질의 종류를 구별하는 문제이다.

핵심 개념

이 방법은 문서의 각 문장이 요약문에 포함될 가능성을 문장 자체의 길이와 표현, 문서 안의 위치, 주제 관련 단어 등의 자질로 계산한다. 대표 자질은 문장길이, 고정 단어구, 문단·문장 위치, 주제어, 대문자 단어이며, 단어빈도 자체는 독립된 문장 추출 자질 유형으로 제시되지 않는다.

선지별 해설

① O — 문장길이는 지나치게 짧은 문장처럼 요약문에 포함될 가능성이 낮은 문장을 판별하는 자질이다.
② O — 고정 단어구는 결론이나 요지를 드러내는 단서 표현이 문장에 포함되어 있는지를 나타내는 자질이다.
③ O — 문단/문장 위치는 문장이 문서나 문단의 어느 위치에 놓였는지를 반영하는 자질이다.
④ X — 단어빈도 자체는 제시된 독립 자질 유형이 아니다. 틀린 표현인 '단어빈도'는 대표 자질 가운데 하나인 '대문자 단어'로 고쳐야 한다. 단어의 빈도는 주제어를 정하는 과정에 활용될 수 있지만, 주제어 자질과 별개의 추출 자질로 열거된 것은 아니다.
⑤ O — 주제어는 문서의 핵심 내용을 나타내는 단어가 해당 문장에 얼마나 포함되어 있는지를 반영하는 자질이다.

오답 포인트

주제어를 선정할 때 단어의 출현 빈도를 활용할 수 있다는 점 때문에 ④ 단어빈도를 별도의 문장 추출 자질로 착각하기 쉽다. 이 문항은 주제어를 만드는 데 쓰일 수 있는 정보와 최종적으로 분류기에 입력되는 자질 유형을 구분해야 한다.

NoteBase

다시 나오면 이렇게 판단

문장 단위 자질인지 먼저 확인한다. 문장길이와 고정 단어구는 문장 자체의 형식·표현, 문단/문장 위치는 구조, 주제어와 대문자 단어는 내용 단서에 해당한다. 단어빈도는 주제어 산출의 바탕이 될 수 있으나 독립 자질 목록에는 포함되지 않는다.

암기 포인트

길이·고정구·위치·주제어·대문자 단어는 자질이고, 단어빈도는 독립 자질이 아니다.

NoteBase

11. 자동색인에서 적합한 색인어를 선정하기 위한 기법 중 성격이 다른 하나는?

① 확률색인 기법
② 구문분석 기법
③ 의미분석 기법
④ 단서어 기법
⑤ 문헌구조적 기법

정답 1번

출제 포인트

자동색인의 색인어 선정 기법을 통계적 기법과 비통계적 기법으로 구분할 수 있는지를 묻는 문제이다.

핵심 개념

확률색인 기법은 단어가 주제어일 가능성이나 적합성 등의 확률을 계산하여 색인어를 선정하는 통계적 기법이다. 반면 구문분석·의미분석·단어·문헌구조적 기법은 단어와 문장의 언어적 특성 또는 문헌 내 위치와 구조를 이용하는 비통계적 기법이다. 결정적인 구분 기준은 확률값을 계산하여 선정하는가이다.

선지별 해설

① O — 확률색인 기법은 확률 모형과 통계값을 이용해 색인어의 중요성이나 적합성을 판단하는 통계적 기법이므로 나머지와 성격이 다르다.
② X — 구문분석 기법은 문장 속 단어와 단어구의 문법적·통사적 관계를 분석하여 색인어를 고르는 비통계적 기법이다. 성격이 다른 기법으로 분류하는 것은 틀리며, 올바르게는 비통계적·언어학적 기법으로 본다.
③ X — 의미분석 기법은 문맥에서 단어가 나타내는 의미를 분석하여 대표 개념을 색인어로 선정하는 비통계적 기법이다. 성격이 다른 기법으로 분류하는 것은 틀리며, 올바르게는 비통계적·언어학적 기법으로 본다.
④ X — 단어 기법은 문헌에 나타난 단어 자체를 대상으로 색인어 후보를 추출·선정하는 비통계적 기법이다. 성격이 다른 기법으로 분류하는 것은 틀리며, 올바르게는 언어적 요소를 이용하는 비통계적 기법으로 본다.
⑤ X — 문헌구조적 기법은 표제·초록·절 제목 등 문헌의 구조와 단어가 나타난 위치를 이용하여 색인어를 선정하는 비통계적 기법이다. 성격이 다른 기법으로 분류하는 것은 틀리며, 올바르게는 문헌의 구조적 특성을 이용하는 비통계적 기법으로 본다.

오답 포인트

구문분석·의미분석만 비통계적 기법으로 보고 문헌구조적 기법을 별개의 성격으로 오인하기 쉽다. 이 문항의 큰 구분은 확률 계산을 사용하는 ①과 언어·문헌의 특성을 이용하는 ②~⑤의 대비이다.

NoteBase

다시 나오면 이렇게 판단

선지에 확률, 빈도, 가중치처럼 수치 계산을 핵심으로 하는 표현이 있으면 통계적 기법으로 판단한다. 구문, 의미, 단어 자체, 문헌의 위치·구조를 분석하면 비통계적 기법으로 판단한다.

암기 포인트

확률은 통계, 구문·의미·단어·문헌구조는 비통계이다.

NoteBase

12. <보기>의 법칙을 제시한 정보학자에 대한 설명으로 옳은 것은?

<보 기>

정보검색시스템은 이용하는 것보다 이용하지 않는 것이 더 문제가 될 때에만 이용될 것이다.

① 조합색인을 창안하였다.
② SDI 시스템을 개발하였다.
③ 정보검색이라는 용어를 처음 사용하였다.
④ 불리언 논리를 색인과 검색에 도입하였다.
⑤ IBM의 정보검색 연구 관리자였다.

정답 3번

출제 포인트

이 문제는 무어스의 법칙을 제시한 캘빈 무어스와 정보검색 분야의 주요 학자·업적을 연결하는 문제이다.

핵심 정리

캘빈 무어스: 정보검색 시스템이 유용하더라도 정보를 얻고 활용하는 일이 더 번거롭다고 느껴지면 이용되지 않는다는 '무어스의 법칙'을 제시했고, '정보검색'이라는 용어를 처음 사용하였다.
모티머 타우브: 유니텀에 기초한 조합색인을 창안하고 불리언 논리를 색인과 검색에 도입하였다.
한스 피터 룬: IBM에서 정보검색 연구를 이끌었으며 SDI 시스템, 자동색인, KWIC 색인 등의 발전에 기여하였다.

선지별 해설

① X — '조합색인을 창안하였다'는 무어스의 업적이 아니다. 올바르게는 모티머 타우브가 유니텀에 기초한 조합색인을 창안하였다.
② X — 'SDI 시스템을 개발하였다'는 무어스의 업적이 아니다. 올바르게는 한스 피터 룬이 이용자의 관심 분야에 맞추어 새 정보를 선별 제공하는 SDI 시스템을 개발하였다.
③ O — 캘빈 무어스는 '정보검색(information retrieval)'이라는 용어를 처음 사용한 학자이다. 제시된 법칙도 그의 이름을 딴 무어스의 법칙이므로 옳다.
④ X — '불리언 논리를 색인과 검색에 도입하였다'는 무어스의 업적이 아니다. 올바르게는 모티머 타우브가 조합색인에 불리언 논리를 적용하여 검색어의 논리적 결합을 가능하게 하였다.
⑤ X — 'IBM의 정보검색 연구 관리자였다'는 무어스에 대한 설명이 아니다. 올바르게는 한스 피터 룬이 IBM에서 정보검색 연구를 이끈 인물이다.

NoteBase

출제자가 노린 함정

무어스, 타우브, 룬은 모두 초기 정보검색 발전에 기여했기 때문에 업적이 뒤섞이기 쉽다. 특히 ③의 '정보검색이라는 용어'는 무어스, ①·④의 '조합색인·불리언 논리'는 타우브, ②·⑤의 'SDI·IBM'은 룬과 연결해야 한다.

다시 나오면 이렇게 판단

보기의 핵심 문장이 '정보를 얻는 고통이 얻지 않는 것보다 크면 시스템을 이용하지 않는다'는 내용이면 무어스의 법칙으로 판단한다. 이어서 무어스는 '정보검색 용어', 타우브는 '조합색인·불리언 논리', 룬은 'SDI·IBM'으로 구분한다.

암기 포인트

무어스=정보검색 용어·무어스의 법칙, 타우브=조합색인·불리언 논리, 룬=SDI·IBM이다.

NoteBase

13. <보기>에 제시된 정보탐색기법에 대한 설명으로 옳은 것은?

<보 기>

일부를 취해서 보거나, 훑어보는 등의 행동을 통해 정보를 찾고 선택하는 방법으로 특별한 훈련을 필요로 하지 않고 누구나 쉽게 할 수 있는 정보추구 행위이다.

① 비공식적이며, 발견적인 전략이다.
② 이용자의 인지적 부하가 탐색보다 적다.
③ 정보를 찾는 데 걸리는 시간이 비교적 적다.
④ 이용자로 하여금 조직적으로 정보를 찾도록 한다.
⑤ 정독하기, 관찰하기, 항해하기, 모니터하기의 4가지 전략이 있다.

정답 1번

출제 포인트

보기의 정보추구 행위가 브라우징임을 파악하고, 브라우징의 비공식성·발견성·시간 비용과 세부 전략을 구분하는 문제이다.

핵심 개념

브라우징은 정보를 정교한 질의로 검색하기보다 자료의 일부를 훑어보고 선택하면서 관련 정보를 발견하는 상호작용적 정보추구 방식이다. 비공식적·발견적·기회주의적 성격이 강하며, 특별한 검색 훈련 없이도 사용할 수 있다. 분석적 탐색은 미리 세운 검색어와 절차에 따라 목표 정보를 체계적으로 찾는 방식이라는 점에서 브라우징과 구별된다.

NoteBase

선지별 해설

① O — 브라우징은 정형화된 질의나 전문적인 탐색 절차보다 훑어보기와 인식에 의존하므로 비공식적이며, 탐색 과정에서 우연히 관련 정보를 발견하는 발견적 전략이다.

② X — 틀린 부분은 '인지적 부하가 탐색보다 적다'고 일률적으로 단정한 표현이다. 브라우징은 분석적 탐색보다 사전 질의 구성의 부담이 낮을 수 있지만, 불명확하거나 복잡한 대상을 계속 살피고 관련성을 판단할 때에는 인지적 부하가 커질 수 있다. 올바르게는 인지적 부하는 브라우징의 유형과 과제의 복잡성에 따라 달라진다고 해야 한다.

③ X — 틀린 부분은 '정보를 찾는 데 걸리는 시간이 비교적 적다'이다. 브라우징은 여러 정보 항목을 차례로 살펴보고 그때그때 관련성을 판단하므로, 목표가 분명하고 검색 도구를 사용할 수 있는 분석적 탐색보다 시간이 더 걸릴 수 있다.

④ X — 틀린 부분은 '이용자로 하여금 조직적으로 정보를 찾도록 한다'이다. 브라우징은 정보환경의 단서에 반응하며 진행되는 비공식적·기회주의적 탐색이다. 올바르게는 이용자가 반드시 조직적인 절차를 따르는 것이 아니라, 훑어보기와 선택을 반복하면서 정보를 발견한다고 해야 한다.

⑤ X — 틀린 부분은 '정독하기'이다. 브라우징의 네 전략은 훑어보기, 관찰하기, 항해하기, 모니터하기이며, 내용을 자세히 읽는 정독은 훑어보기와 다르다.

오답 포인트

②는 브라우징이 질의 작성 부담을 줄일 수 있다는 점을 '모든 탐색보다 인지적 부하가 작다'는 절대적 특징으로 확대한다. ⑤는 '훑어보기'를 그와 반대되는 '정독하기'로 바꾸어 놓은 것이 핵심 함정이다.

다시 나오면 이렇게 판단

'일부를 훑어봄', '선택하면서 발견함', '특별한 훈련이 필요 없음'이 제시되면 브라우징으로 판단한다. 브라우징의 기본 성격은 비공식적·발견적이며, 시간이나 인지적 부하는 과제에 따라 달라진다는 기준으로 선지를 가른다.

암기 포인트

브라우징은 비공식적·발견적 정보추구이며, 네 전략은 훑어보기·관찰하기·항해하기·모니터하기이다.

NoteBase

14. 다음 중 OAI(Open Archives Initiative) 프로토콜에 대한 설명으로 옳은 것은?

① 인터넷상의 다양한 형태의 물리적 네트워크에 연결된 컴퓨터 간 통신을 위한 절차와 메시지 교환 규칙을 정의한 규약이다.

② 네트워크상의 서로 다른 기종의 컴퓨터 사이에서 정보검색의 상호운용성을 지원하기 위해 1988년 미국표준협회가 채택한 표준 규약이다.

③ 저장소(repository) 간 메타데이터의 효율적인 공유와 활용을 목적으로 상호운용성에 입각하여 개발된 국제표준 통신규약이다.

④ 단순객체접근 프로토콜로, XML과 HTTP 등을 기반으로 하여 다른 컴퓨터에 있는 데이터나 서비스를 호출하기 위한 통신규약이다.

⑤ 웹상에서 주로 HTML 문서를 주고받는 데 사용하는 통신규약이다.

정답 3번

출제 포인트

OAI-PMH의 목적과 작동 방식을 다른 네트워크·정보검색 프로토콜과 구별하는 문제이다.

핵심 개념

OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting)는 저장소가 보유한 메타데이터를 외부에서 수집할 수 있도록 공개하여 저장소 간 상호운용성을 높이는 프로토콜이다. HTTP를 통해 요청과 응답을 주고받고, 메타데이터는 XML 형식으로 전달한다.

TCP/IP는 인터넷에서 컴퓨터 간 통신을 가능하게 하는 프로토콜 체계이고, Z39.50은 서로 다른 정보검색 시스템 사이의 검색·결과 전송을 지원하는 프로토콜이다. SOAP은 XML 기반 메시지로 네트워크상의 서비스와 정보를 교환하는 프로토콜이며, HTTP는 웹에서 문서 등 자원을 요청하고 전송하는 프로토콜이다.

NoteBase

선지별 해설

① X — '물리적 네트워크에 연결된 컴퓨터 간 통신을 위한 절차와 메시지 교환 규칙'은 OAI-PMH가 아니라 TCP/IP 계열 프로토콜에 관한 설명이다. 올바르게는 OAI-PMH는 저장소의 메타데이터를 수집·공유하기 위한 규약이다.
② X — '서로 다른 기종의 컴퓨터 사이에서 정보검색의 상호운용성을 지원'하는 표준은 OAI-PMH가 아니라 Z39.50이다. 올바르게는 OAI-PMH가 저장소의 메타데이터 수집을 통한 상호운용성을 지원한다고 해야 한다.
③ O — OAI-PMH는 저장소가 메타데이터를 효율적으로 공개하고 수집하게 함으로써 메타데이터 공유와 활용, 저장소 간 상호운용성을 지원한다.
④ X — 'XML과 HTTP 등을 기반으로 다른 컴퓨터의 데이터나 서비스를 호출'하는 것은 SOAP에 관한 설명이다. 올바르게는 OAI-PMH가 HTTP 요청과 XML 형식의 응답을 이용하여 저장소의 메타데이터를 수집하는 프로토콜이라고 해야 한다.
⑤ X — '웹상에서 주로 HTML 문서를 주고받는 데 사용하는 통신규약'은 HTTP이다. 올바르게는 OAI-PMH가 웹 환경에서 저장소의 메타데이터를 수집·공유하는 데 사용되는 프로토콜이라고 해야 한다.

오답 포인트

④에는 OAI-PMH도 사용하는 XML과 HTTP가 제시되어 있어 혼동하기 쉽다. 그러나 OAI-PMH의 목적은 메타데이터 수집이고, SOAP의 목적은 XML 메시지를 통한 서비스 호출과 정보 교환이다. 기반 기술이 같아도 프로토콜의 목적을 기준으로 구별해야 한다.

다시 나오면 이렇게 판단

선지에서 '저장소', '메타데이터', '수집(harvesting)', '상호운용성'이 함께 나타나면 OAI-PMH로 판단한다. 컴퓨터 간 기본 통신은 TCP/IP, 분산 정보검색은 Z39.50, XML 기반 서비스 호출은 SOAP, 웹 문서 전송은 HTTP로 구별한다.

암기 포인트

OAI-PMH는 저장소의 메타데이터를 수확(harvesting)하는 프로토콜이다.

NoteBase

15. 정보검색시스템의 평가를 위하여 널리 사용되고 있는 '사용성'의 속성 중 닐슨(Nielsen)이 제시한 것과 거리가 먼 것은?

① 학습성(learnability): 처음 디자인을 접했을 때 이용자들이 기본 업무를 수행하기가 얼마나 쉬운가?
② 효율성(efficiency): 일단 디자인을 배우고 나면 이용자들이 얼마나 빨리 업무를 수행할 수 있는가?
③ 융통성(flexibility): 시스템이 이용자들의 활용능력 수준을 얼마나 잘 반영할 수 있는가?
④ 기억용이성(memorability): 이용자가 한동안 사용하지 않다가 다시 디자인을 접했을 때 얼마나 쉽게 능숙해질 수 있는가?
⑤ 오류(errors): 이용자들이 얼마나 많은 실수를 하는가?

정답 3번

출제 포인트

Nielsen이 제시한 사용성의 다섯 속성인 학습성, 효율성, 기억용이성, 오류, 만족도를 구별하는 문제이다.

핵심 개념

사용성은 사용자가 시스템을 얼마나 쉽게 배우고, 효율적으로 사용하며, 다시 사용할 때 쉽게 기억하고, 오류를 적게 범하면서 만족스럽게 이용할 수 있는지를 나타낸다. Nielsen의 다섯 속성은 학습성(처음에도 쉽게 배움), 효율성(배운 뒤 빠르게 수행함), 기억용이성(쉬었다가 다시 써도 쉽게 익숙해짐), 오류(실수가 적고 복구가 쉬움), 만족도(주관적으로 만족스럽게 사용함)이다. 융통성 또는 유연성은 이 다섯 속성에 포함되지 않는다.

선지별 해설

① O — 학습성은 처음 시스템이나 디자인을 접한 사용자가 기본 업무를 얼마나 쉽게 수행할 수 있는지를 뜻한다. 제시된 설명은 학습성에 해당한다.
② O — 효율성은 사용법을 익힌 뒤 사용자가 업무를 얼마나 빠르게 수행할 수 있는지를 뜻한다. 제시된 설명은 효율성에 해당한다.
③ X — '융통성(flexibility)'은 Nielsen이 제시한 다섯 가지 사용성 속성이 아니다. 올바르게는 학습성, 효율성, 기억용이성, 오류, 만족도 가운데 하나가 제시되어야 하며, 이 선지에는 빠진 속성인 만족도, 즉 시스템 사용이 얼마나 즐겁고 만족스러운가가 들어갈 수 있다.
④ O — 기억용이성은 일정 기간 사용하지 않은 뒤 다시 접했을 때 사용법을 얼마나 쉽게 되살려 능숙하게 사용할 수 있는지를 뜻한다. 제시된 설명은 기억용이성에 해당한다.
⑤ O — 오류는 사용자가 얼마나 적게 실수하는지, 오류에서 얼마나 쉽게 회복하는지, 치명적인 오류가 발생하지 않는지를 살피는 속성이다. '얼마나 많은 실수를 하는가'라는 질문은 오류의 빈도를 평가하므로 오류 속성에 해당한다.

NoteBase

오답 포인트

③의 '이용자들의 활용능력 수준을 얼마나 잘 반영하는가'라는 설명이 사용성의 일반적 의미처럼 보여 혼동하기 쉽다. 그러나 이 문항은 사용성과 관련 있어 보이는지를 묻는 것이 아니라 Nielsen이 정리한 다섯 속성에 실제로 포함되는지를 묻는다. 또한 ⑤는 오류의 모든 평가 요소를 열거하지 않았지만, 실수의 수를 묻는 것만으로도 오류 속성에 해당한다.

다시 나오면 이렇게 판단

Nielsen의 사용성 속성 문항에서는 '처음-배운 뒤-오랜만에-실수-느낌'의 순서로 대응한다. 처음은 학습성, 배운 뒤의 속도는 효율성, 오랜만에 다시 쓰는 것은 기억용이성, 실수와 복구는 오류, 주관적 즐거움은 만족도이다. 이 다섯 범주에 없는 용어를 고르면 된다.

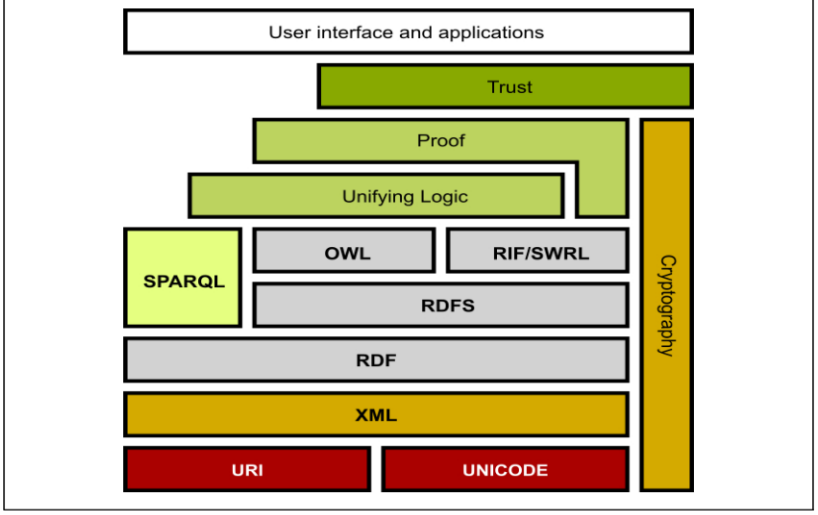
암기 포인트

Nielsen의 사용성 5속성은 '학습-효율-기억-오류-만족'이다.

NoteBase

16. <보기>는 시맨틱 웹 계층구조(Semantic Web Stack)에 대한 그림이다. 각 구성요소를 설명한 것으로 옳지 않은 것은?

<보 기>



① URI - 정보자원의 주소를 지정하는 방법 제공

② UNICODE - 전 세계의 모든 문자를 컴퓨터에서 일관되게 표현하기 위한 문자 집합 표준

③ XML - 자원의 의미정보를 표현할 수 있는 기반 제공

④ RDF - 웹에 있는 자원에 관한 메타정보를 표현하기 위해 웹컨소시엄(W3C)이 권고하는 언어

⑤ OWL - 관계형 데이터베이스의 데이터에 접근하기 위한 쿼리 언어로 국제표준화기구(ISO)가 권고하는 언어

정답 5번

출제 포인트

시맨틱 웹 계층구조의 기초 요소인 URI·UNICODE·XML·RDF와 온톨로지 언어 OWL의 역할을 구별하는 문제이다.

핵심 개념

URI는 웹 자원을 식별하는 주소 체계이고, UNICODE는 세계 여러 문자를 일관되게 표현하는 문자 부호 표준이다. XML은 태그를 이용해 데이터를 구조화하는 구문적 기반이며, RDF는 자원에 관한 정보를 주어-술어-목적어 형태로 표현하는 데이터 모델이다. OWL은 개념과 개념 사이의 관계 및 제약을 정의하는 웹 온톨로지 언어이다.

NoteBase

선지별 해설

① O — URI는 웹상의 정보 자원을 고유하게 식별하고 그 위치나 이름을 나타내는 주소 지정의 기반을 제공한다.
② O — UNICODE는 언어와 운영체제에 관계없이 세계 여러 문자를 일관된 코드 체계로 표현하기 위한 문자 집합 표준이다.
③ O — XML은 태그를 사용해 데이터를 구조화하므로 시맨틱 웹에서 의미 정보를 기술·교환할 수 있는 구문적 기반을 제공한다. 다만 의미와 관계 자체를 직접 모델링하는 핵심 역할은 RDF와 OWL이 담당한다.
④ O — RDF는 웹 자원에 관한 메타데이터와 자원 간 관계를 표현하기 위해 W3C가 표준화한 언어이자 데이터 모델이다.
⑤ X — 틀린 표현은 '관계형 데이터베이스의 데이터에 접근하기 위한 쿼리 언어'와 'ISO가 권고하는 언어'이다. OWL은 W3C가 표준화한 웹 온톨로지 언어로, 개념·관계·제약을 명시적으로 정의한다. 관계형 데이터베이스의 대표적인 질의 언어는 SQL이며, RDF 데이터 질의에는 SPARQL을 사용한다.

오답 포인트

⑤는 시맨틱 웹 구성요소인 OWL에 다른 기술의 역할을 섞어 놓았다. '쿼리 언어'라는 표현은 SPARQL 또는 SQL의 역할이고, OWL은 온톨로지 기술 언어이다. 또한 OWL의 표준화 기구는 ISO가 아니라 W3C이다.

다시 나오면 이렇게 판단

각 계층을 '식별-문자-구조-자원 관계-온톨로지-질의'로 나누어 판단한다. URI는 식별, UNICODE는 문자, XML은 구조화 구문, RDF는 자원과 관계의 표현, OWL은 온톨로지, SPARQL은 RDF 질의를 담당한다.

암기 포인트

URI는 주소, UNICODE는 문자, XML은 구조, RDF는 자원 관계, OWL은 온톨로지, SPARQL은 질의이다.

NoteBase

17. 디지털 아카이빙의 문제점 중 저장매체의 물리적 약화를 보완할 수 있는 방법을 <보기>에서 모두 고른 것은?

<보 기>

ㄱ. 복사	ㄴ. 인캡슐레이션
ㄷ. 마이그레이션	ㄹ. 에뮬레이션

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

정답 2번

출제 포인트

디지털 아카이빙에서 저장매체의 물리적 열화를 직접 보완하는 방법과, 파일의 이용 환경을 보존하는 방법을 구별하는 문제이다.

핵심 개념

복사: 디지털 정보를 다른 저장매체에 다시 기록하여 손상·열화 위험에 대비하는 방법이다.
마이그레이션: 디지털 정보를 새로운 저장매체나 시스템 환경으로 이전하여 기존 매체의 열화와 노후화에 대응하는 방법이다.
인캡슐레이션: 디지털 객체와 함께 그 객체를 해석하는 데 필요한 메타데이터·관련 정보를 묶어 보존하는 방법이다.
에뮬레이션: 구형 하드웨어나 소프트웨어 환경을 현재 시스템에서 모방하여 원래 디지털 객체를 이용하게 하는 방법이다.
따라서 매체 자체의 물리적 약화를 보완하는 것은 ㄱ 복사와 ㄷ 마이그레이션이다.

NoteBase

선지별 해설

① X — ㄱ 복사는 해당하지만 ㄴ 인캡슐레이션은 매체의 물리적 약화를 직접 보완하는 방법이 아니다. 틀린 구성은 'ㄱ, ㄴ'이며, 올바르게는 'ㄱ, ㄷ'이다.
② O — ㄱ 복사와 ㄷ 마이그레이션은 정보를 다른 매체에 다시 기록하거나 새로운 매체·환경으로 이전하므로 저장매체의 물리적 열화에 대응할 수 있다.
③ X — ㄴ 인캡슐레이션과 ㄹ 에뮬레이션은 디지털 객체의 해석 정보나 이용 환경을 보존하는 방법이지, 저장매체의 물리적 약화를 직접 보완하는 방법이 아니다. 틀린 구성은 'ㄴ, ㄹ'이며, 올바르게는 'ㄱ, ㄷ'이다.
④ X — ㄱ 복사와 ㄷ 마이그레이션은 해당하지만 ㄴ 인캡슐레이션이 불필요하게 포함되었다. 틀린 구성은 'ㄱ, ㄴ, ㄷ'이며, 올바르게는 'ㄱ, ㄷ'이다.
⑤ X — ㄱ 복사는 해당하지만 ㄴ 인캡슐레이션과 ㄹ 에뮬레이션은 해당하지 않고, 필요한 ㄷ 마이그레이션이 빠졌다. 틀린 구성은 'ㄱ, ㄴ, ㄹ'이며, 올바르게는 'ㄱ, ㄷ'이다.

오답 포인트

인캡슐레이션과 에뮬레이션도 디지털 보존 기법이라는 점 때문에 모두 물리적 열화 대책으로 고르기 쉽다. 그러나 이 문항의 핵심 표현은 '저장매체의 물리적 약화'이다. 인캡슐레이션은 해석에 필요한 정보를 함께 보존하고, 에뮬레이션은 구형 실행 환경을 재현한다는 점에서 매체 자체를 새로 기록하거나 교체하는 방법과 다르다.

다시 나오면 이렇게 판단

문제에서 저장매체의 손상·열화가 제시되면 정보가 실제로 다른 매체에 다시 기록되는지를 판단한다. 다시 기록하거나 이전하는 복사·마이그레이션은 해당하고, 정보 묶음이나 실행 환경을 보존하는 인캡슐레이션·에뮬레이션은 제외한다.

암기 포인트

매체의 물리적 열화에는 복사·마이그레이션, 해석 정보에는 인캡슐레이션, 실행 환경에는 에뮬레이션이다.

NoteBase

18. 원문(full-text)에 대한 검색도구로서의 문헌대체물로 색인이나 범주화, 요약과 같은 정보표현 방법이 사용된다. 다음 중 Meadow가 제시한 정보표현을 위한 특정 방법의 선택 시 고려해야 할 사항이 아닌 것은?

① 엔티티에 대한 정확한 서술 가능성
② 유사한 엔티티의 확인
③ 해석의 모호성 최소화
④ 엔티티의 망라성
⑤ 다른 엔티티 간의 구별

정답 4번

출제 포인트

이 문제는 원문을 대신하는 정보표현 방법을 선택할 때 표현의 정확성·식별성·명확성을 평가하는 기준과, 그 기준에 해당하지 않는 요소를 구별하는 문제이다.

핵심 개념

정보표현 방법은 원문의 엔터티를 정확하게 나타내고, 서로 비슷한 엔터티는 관련성을 확인할 수 있게 하며, 서로 다른 엔터티는 구별할 수 있게 해야 한다. 또한 표현을 여러 뜻으로 해석할 가능성을 줄여야 한다. 반면 엔터티의 망라성은 대상 엔터티를 빠짐없이 포괄하는 정도로서, 특정 정보표현 방법 자체의 선택 기준과는 구별된다.

선지별 해설

① O — 정보표현은 대상 엔터티의 내용을 정확하게 서술할 수 있어야 하므로 고려 사항에 해당한다.
② O — 유사한 엔터티의 관련성이나 유사성을 확인할 수 있어야 검색 결과를 적절히 묶고 비교할 수 있으므로 고려 사항에 해당한다.
③ O — 같은 표현이 여러 의미로 해석되는 모호성을 줄여야 일관된 색인과 검색이 가능하므로 고려 사항에 해당한다.
④ X — '엔터티의 망라성'은 대상 엔터티를 모두 포괄하는가에 관한 속성으로, 특정 정보표현 방법을 선택할 때의 표현 품질 기준이 아니다. 올바르게는 정보표현 방법이 엔터티를 정확하고 명확하게 나타내며 유사한 엔터티를 확인하고 서로 다른 엔터티를 구별할 수 있는지를 고려한다.
⑤ O — 서로 다른 엔터티가 같은 것으로 처리되지 않도록 구별할 수 있어야 하므로 고려 사항에 해당한다.

오답 포인트

④의 '망라성'도 좋은 검색도구의 일반적 장점처럼 보여 혼동하기 쉽다. 그러나 이 문항은 원문을 대신할 특정 정보표현 방법의 선택 기준을 묻는다. 정확한 서술, 유사성 확인, 모호성 최소화, 상이한 엔터티의 구별은 표현의 기능이지만, 엔터티 전체를 빠짐없이 포괄하는 망라성은 그와 다른 차원의 속성이다.

NoteBase

다시 나오면 이렇게 판단

선지가 '한 엔터티를 얼마나 정확하고 명확하게 표현하며 다른 엔터티와 식별하게 하는가'를 묻는다면 정보표현 방법의 선택 기준으로 판단한다. 반대로 '대상 전체를 빠짐없이 포함하는가'를 묻는다면 표현 방법의 품질이 아니라 범위의 망라성에 관한 항목으로 구별한다.

암기 포인트

정보표현은 정확성·유사성 확인·모호성 최소화·구별성이고, 망라성은 별개의 범위 문제이다.

NoteBase

19. W3C의 디렉터 팀 버너스리(Tim Berners-Lee)가 서술한 개방형연결데이터(Linked Open Data, LOD)의 네 가지 원칙으로 옳은 것은?

① URI로 식별하여 연결하고, RDF로 표현하고, SQL로 질의하고, FTP로 유통한다.

② URI로 식별하여 연결하고, RDF로 표현하고, SPARQL로 질의하고, HTTP로 유통한다.

③ URI로 식별하여 연결하고, XML로 표현하고, SQL로 질의하고, TCP/IP로 유통한다.

④ URI로 식별하여 연결하고, HTML로 표현하고, SQL로 질의하고, HTTP로 유통한다.

⑤ URI로 식별하여 연결하고, HTML로 표현하고, SPARQL로 질의하고, HTTP로 유통한다.

정답 2번

출제 포인트

개방형연결데이터(LOD)의 네 가지 원칙에 대응하는 기술인 URI, RDF, SPARQL, HTTP를 정확히 구별하는 문제이다.

핵심 개념

LOD는 웹에서 데이터를 식별하고 공개하며 서로 연결해 기계가 의미를 이해하고 활용할 수 있게 한 데이터이다. 각 자원은 URI로 식별하고 다른 자원과 연결하며, 데이터는 RDF로 표현하고 SPARQL로 질의하며, HTTP를 통해 접근·유통한다.

RDF는 데이터를 주어-술어-목적어 구조로 표현하는 표준이고, SPARQL은 RDF 데이터를 검색하는 질의 언어이다. SQL은 주로 관계형 데이터베이스의 질의 언어이며, HTML은 웹 문서의 표현 언어이므로 RDF와 역할이 다르다.

선지별 해설

① X — SQL과 FTP가 잘못되었다. 올바르게는 URI로 식별·연결하고, RDF로 표현하며, SPARQL로 질의하고, HTTP로 유통한다.

② O — URI, RDF, SPARQL, HTTP가 각각 자원의 식별·연결, 데이터 표현, 질의, 웹을 통한 접근·유통에 대응하므로 옳다.

③ X — XML, SQL, TCP/IP가 LOD의 원칙에 대응하는 기술로 제시된 부분이 잘못되었다. 올바르게는 RDF로 표현하고 SPARQL로 질의하며 HTTP로 유통한다.

④ X — HTML과 SQL이 잘못되었다. 올바르게는 RDF로 표현하고 SPARQL로 질의하며, HTTP로 유통한다.

⑤ X — HTML이 잘못되었다. 올바르게는 RDF로 표현하며, 나머지는 SPARQL로 질의하고 HTTP로 유통한다.

NoteBase

오답 포인트

웹과 데이터베이스에서 널리 쓰이는 XML·HTML·SQL·FTP·TCP/IP를 모두 LOD의 핵심 기술로 착각하기 쉽다. 특히 SQL은 관계형 데이터베이스용 질의 언어이고, LOD의 RDF 데이터에는 SPARQL을 사용한다는 차이가 핵심 함정이다.

다시 나오면 이렇게 판단

선지에서 식별·연결은 URI, 표현은 RDF, 질의는 SPARQL, 접근·유통은 HTTP인지 순서대로 대응시킨다. 이 네 역할 중 하나라도 HTML·XML·SQL·FTP 등으로 바뀌면 틀린 선지로 판단한다.

암기 포인트

LOD는 ‘URI로 찾고, RDF로 쓰고, SPARQL로 묻고, HTTP로 연다’로 정리된다.

NoteBase

20. <보기>는 정보추구행태에서 나타나는 행동들을 단계화 한 쿨타우(Carol C. Kuhlthau)의 정보탐색과정(Information Search Process) 모형의 각 단계를 설명한 것이다. 각 단계와 해당 명칭을 짝지은 것으로 옳지 않은 것은?

<보 기>

가. 정보추구자가 어떤 지식이나 이해가 필요하다고 인식하는 불확실성 단계

나. 정보추구자는 특정한 주제 초점을 맞추면서, 그 정보요구를 충족시키기 위한 최선의 접근방법을 발견하는 단계

다. 정보추구자가 주제 방향을 제공하는 정보를 수집하면서, 주제에 대한 면밀한 탐구를 시작하는 단계

라. 정보추구자가 탐구를 위한 명백한 초점을 확립하고 불확실성이 줄어들기 시작하는 단계

마. 정보추구자는 정보의 수집과정에서 초점이 명확해지고, 확정된 주제와 관련된 정보만을 수집하는 단계

① 가 - 과제 착수(task initiation)
② 나 - 주제 선정(topic selection)
③ 다 - 주제 탐구(topic exploration)
④ 라 - 정보 선택(making selection)
⑤ 마 - 정보 수집(resources collection)

정답 4번

출제 포인트

쿨타우의 정보탐색과정에서 나타나는 행동의 특징과 과제 착수·주제 선정·주제 탐구·초점 형성·정보 수집 단계를 정확히 연결하는 문제이다.

핵심 정리

과제 착수(task initiation): 지식이나 이해가 부족하여 정보가 필요함을 처음 인식하는 단계이다.
주제 선정(topic selection): 탐색할 주제와 정보요구를 충족할 기본 접근방법을 선택하는 단계이다.
주제 탐구(topic exploration): 주제 전반의 정보를 살펴 이해를 넓히고 탐색 방향을 구체화하는 단계이다.
초점 형성(focus formulation): 탐구 결과를 바탕으로 명확한 관점이나 초점을 세우며 불확실성이 감소하는 단계이다.
정보 수집(information collection): 형성된 초점과 직접 관련된 정보를 집중적으로 모으는 단계이다.

NoteBase

선지별 해설

① O — 가는 정보추구자가 지식이나 이해의 부족을 깨닫고 정보의 필요성을 인식하는 과제 착수 단계이므로 연결이 옳다.
② O — 나는 탐색할 특정 주제에 초점을 맞추고 정보요구를 충족할 접근방법을 찾는 주제 선정 단계이므로 연결이 옳다.
③ O — 다는 주제 방향을 알려 주는 일반적 정보를 수집하면서 주제에 대한 이해를 넓히는 주제 탐구 단계이므로 연결이 옳다.
④ X — 라의 '명백한 초점을 확립하고 불확실성이 줄어들기 시작한다'는 표현은 정보 선택(making selection)이 아니라 초점 형성(focus formulation)에 해당한다. 올바르게는 '라 - 초점 형성(focus formulation)'이다.
⑤ O — 마는 이미 명확해진 초점에 따라 확정된 주제와 관련된 정보만 집중적으로 모으는 정보 수집 단계이므로 연결이 옳다.

출제자가 노린 함정

주제 선정과 초점 형성을 모두 '선택'과 비슷한 과정으로 혼동하게 만드는 것이 핵심 함정이다. 특히 라의 '명백한 초점 확립'과 '불확실성 감소'는 단순한 정보 선택이 아니라 초점 형성의 결정적 단서이다.

다시 나오면 이렇게 판단

정보 필요를 처음 느낌은 과제 착수, 탐색할 주제를 고름은 주제 선정, 주제 전반을 살펴봄은 주제 탐구, 관점과 초점이 명확해짐은 초점 형성, 초점에 맞는 정보만 모음은 정보 수집으로 판단한다.

암기 포인트

필요 인식 → 주제 선정 → 넓게 탐구 → 초점 형성 → 좁게 수집의 흐름으로 기억한다.

NoteBase