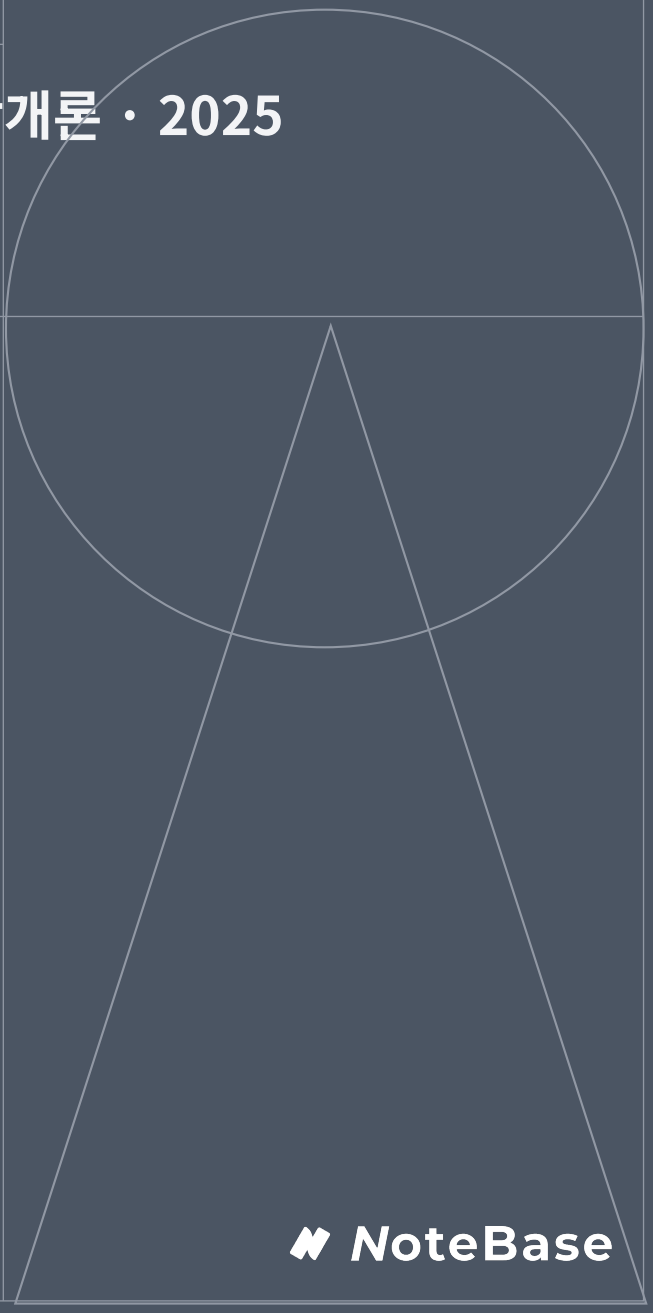
기출OX 해설집

공무원/법원직/9급 · 정보학개론 · 2025



NoteBase

기출OX 해설집

공무원/법원직/9급 · 정보학개론 · 2025

공무원/법원직/9급 · 정보학개론 · 2025

문 1. 정보 이용자의 정보 추구 행태 모델 중 Kuhlthau의 ISP(Information Search Process) 모델의 단계 순서가 바르게 나열된 것은?

① 과제착수 → 주제선정 → 탐색활동 → 초점형성 → 정보수집 → 과제완료

② 주제선정 → 과제착수 → 탐색활동 → 정보수집 → 초점형성 → 과제완료

③ 과제착수 → 주제선정 → 탐색활동 → 초점형성 → 과제완료 → 정보수집

④ 주제선정 → 탐색활동 → 과제착수 → 초점형성 → 정보수집 → 과제완료

정답 1번

출제 포인트

이 문제는 Kuhlthau의 정보탐색과정(ISP)이 과제의 시작에서 탐색·초점 설정·수집을 거쳐 완료로 이어지는 순서를 묻는 문제이다.

올바른 순서

1. 과제착수 — 정보가 필요함을 인식하고 탐색 과제를 시작하는 단계이다.
2. 주제선정 — 탐색할 일반적인 주제를 정하는 단계이다.
3. 탐색활동 — 선정한 주제에 관한 정보를 폭넓게 살펴보며 이해를 넓히는 단계이다.
4. 초점형성 — 탐색 결과를 바탕으로 구체적인 관점이나 초점을 정하는 단계이다.
5. 정보수집 — 형성된 초점에 맞는 관련 정보를 집중적으로 모으는 단계이다.
6. 과제완료 — 탐색을 마무리하고 결과를 제시하는 단계이다.

왜 이 순서인가: 먼저 폭넓게 탐색해야 구체적인 초점을 형성할 수 있고, 그 초점이 정해져야 필요한 정보를 선별하여 수집할 수 있기 때문이다.

NoteBase

선지별 해설

① O — 과제착수 → 주제선정 → 탐색활동 → 초점형성 → 정보수집 → 과제완료로 ISP의 여섯 단계가 정확히 배열되어 있다.

② X — '주제선정 → 과제착수'와 '정보수집 → 초점형성'의 선후관계가 뒤바뀌었다. 올바르게는 과제를 먼저 착수한 뒤 주제를 선정하며, 탐색활동 후 초점을 형성한 다음 그 초점에 맞추어 정보를 수집한다. 전체 순서는 과제착수 → 주제선정 → 탐색활동 → 초점형성 → 정보수집 → 과제완료이다.

③ X — '과제완료 → 정보수집'이 잘못되었다. 과제를 완료하려면 필요한 정보의 수집이 먼저 이루어져야 하므로, 올바르게는 초점형성 → 정보수집 → 과제완료이다.

④ X — '주제선정 → 탐색활동 → 과제착수'가 잘못되었다. 정보 필요를 인식하여 과제에 착수한 다음 주제를 선정하고 탐색활동을 진행하므로, 올바르게는 과제착수 → 주제선정 → 탐색활동이다. 전체 순서는 과제착수 → 주제선정 → 탐색활동 → 초점형성 → 정보수집 → 과제완료이다.

오답 포인트

핵심 함정은 탐색활동·초점형성·정보수집의 순서이다. 폭넓은 탐색으로 관점을 잡은 뒤에야 그 관점에 맞는 정보를 집중적으로 수집하므로, 정보수집이 초점형성보다 앞설 수 없다. 또한 과제착수는 주제선정보다 앞이고, 과제완료는 정보수집보다 뒤이다.

다시 나오면 이렇게 판단

ISP 순서형 문제는 '시작 → 주제 결정 → 폭넓은 탐색 → 초점 결정 → 집중 수집 → 완료'의 흐름으로 판단한다. 특히 탐색활동이 초점형성의 전제이고, 초점형성이 정보수집의 전제라는 관계를 확인한다.

암기 포인트

착수·선정·탐색·초점·수집·완료: 먼저 넓게 탐색하고, 초점을 잡은 뒤, 필요한 정보를 모은다.

NoteBase

문 2. 월드 와이드 웹 컨소시엄(World Wide Web Consortium: W3C)의 웹 콘텐츠 접근성 지침(Web Content Accessibility Guidelines: WCAG) 2.2의 원칙에 관한 다음 설명 중 가장 옳지 않은 것은?

① 웹 접근성의 원칙으로 인식의 용이성, 운용의 용이성, 유연성, 견고성을 제시한다.

② 대체 텍스트, 시간 기반 미디어, 적응 가능, 식별 가능의 지침을 제시한다.

③ 가독성, 예측 가능성, 입력 지원 지침을 제공함으로써 이용자 인터페이스의 정보와 운용을 이해하도록 지원한다.

④ 보조공학을 포함하여, 현재와 미래의 이용자 에이전트와의 호환성을 극대화할 것을 제시한다.

정답 1번

출제 포인트

WCAG 2.2의 네 원칙과 각 원칙 아래에 배치된 지침을 구분하는 문제이다.

핵심 개념

WCAG 2.2의 네 원칙은 인식의 용이성(Perceivable), 운용의 용이성(Operable), 이해의 용이성(Understandable), 견고성(Robust)이다.
인식의 용이성에는 대체 텍스트, 시간 기반 미디어, 적응 가능, 식별 가능 지침이 속한다. 이해의 용이성에는 가독성, 예측 가능성, 입력 지원 지침이 속하고, 견고성에는 호환성 지침이 속한다.

선지별 해설

① X — '유연성'은 WCAG 2.2의 네 원칙에 포함되지 않는다. 틀린 표현은 '유연성'이며, 올바르게는 '이해의 용이성'이다.
② O — 대체 텍스트, 시간 기반 미디어, 적응 가능, 식별 가능은 모두 인식의 용이성 아래의 지침이다.
③ O — 가독성, 예측 가능성, 입력 지원은 이해의 용이성 아래의 지침이며, 정보와 이용자 인터페이스의 운용을 이해할 수 있도록 한다.
④ O — 현재와 미래의 이용자 에이전트 및 보조공학과의 호환성을 극대화하는 것은 견고성 원칙 아래의 호환성 지침이다.

오답 포인트

①의 '유연성'은 접근성에 어울리는 일반 표현처럼 보이지만 WCAG 2.2의 공식 원칙명이 아니다. '적응 가능'은 원칙이 아니라 인식의 용이성 아래의 지침이라는 점과도 구분해야 한다.

다시 나오면 이렇게 판단

먼저 네 원칙을 인식·운용·이해·견고로 고정한 뒤, 선지의 용어가 원칙인지 그 아래 지침인지 구분한다.

NoteBase

암기 포인트

WCAG 네 원칙은 '인·운·이·견', 즉 인식·운용·이해·견고이다.

NoteBase

문 3. 초거대 AI의 기술적 한계 중 '부정확성'에 관한 설명으로 가장 옳지 않은 것은?

① 초거대 AI는 인간처럼 맥락과 논리를 완벽히 이해하고 추론하므로, 제공하는 정보는 항상 최신의 사실을 반영하여 신뢰도가 매우 높다.

② 초거대 AI가 때때로 거짓 정보를 마치 사실인 것처럼 제시하는 환각 현상은 확률 기반 생성 방식의 한계로 인해 발생한다.

③ 챗GPT와 같은 초거대 AI는 특정 시점까지의 데이터로 사전 학습되어, 그 이후의 최신 정보에 대해서는 정확한 답변을 제공하기 어렵다.

④ 초거대 AI는 학습된 데이터 내 단어 및 데이터 간의 관계를 파악하여 확률적으로 가장 그럴듯한 답변을 생성하는 방식으로 작동한다.

정답 1번

출제 포인트

초거대 AI가 문장을 생성하는 방식과 그로 인해 나타나는 부정확성의 원인을 이해하는지를 묻는 문제이다.

핵심 개념

초거대 AI는 학습 데이터에서 익힌 단어와 데이터 사이의 관계를 바탕으로 다음에 올 내용을 확률적으로 생성하는 모델이다. 인간처럼 의미와 논리를 완벽히 이해하거나 모든 답의 사실성을 보장하는 체계가 아니므로, 그럴듯하지만 사실과 다른 내용을 제시하는 환각이 발생할 수 있다. 또한 사전 학습 이후의 정보는 별도의 최신 정보 연결이나 추가 학습이 없다면 반영하기 어렵다.

선지별 해설

① X — '맥락과 논리를 완벽히 이해하고 추론하므로', '항상 최신의 사실을 반영하여 신뢰도가 매우 높다'는 표현이 틀렸다. 올바르게는 초거대 AI는 확률적으로 답변을 생성하므로 맥락이나 사실을 잘못 처리할 수 있고, 학습 시점 이후의 최신 정보가 자동으로 항상 반영되는 것도 아니다.
② O — 실제 사실이 아닌 내용을 사실처럼 제시하는 환각은 확률적으로 그럴듯한 출력을 생성하는 방식에서 나타날 수 있는 대표적인 한계이다.
③ O — 사전 학습 데이터에는 시간적 범위가 있으므로, 별도의 최신 정보 연결이나 추가 학습이 없다면 그 이후의 사건이나 변화에 정확히 답하기 어렵다.
④ O — 학습 데이터에서 파악한 단어와 데이터 사이의 관계를 토대로 다음 내용을 확률적으로 예측하여 그럴듯한 답변을 생성한다는 설명이다.

오답 포인트

①은 '완벽히', '항상'이라는 절대적 표현으로 초거대 AI의 이해 능력과 최신성을 과장한다. 자연스럽고 논리적으로 보이는 문장과 사실이 검증된 문장은 서로 다르다는 점이 핵심 함정이다.

NoteBase

다시 나오면 이렇게 판단

초거대 AI의 정확성을 묻는 선지에서는 '확률적 생성', '환각 가능성', '학습 데이터의 시간적 한계'를 기준으로 판단한다. '완벽히 이해한다', '항상 정확하다', '항상 최신이다'와 같은 절대적 표현은 부정확성의 한계와 모순되는지 확인한다.

암기 포인트

초거대 AI는 사실을 보장하는 장치가 아니라 확률적으로 그럴듯한 답을 생성하는 모델이다. 그럴듯함은 정확함과 같지 않다.

NoteBase

문 4. 다음 중 링크드 데이터(Linked Data) 구축의 기본 원칙(내용)으로 가장 옳지 않은 것은?

① 모든 사물에 대해 URI를 사용하라.

② HTTP URI를 사용하여 이름으로 식별될 수 있게 하라.

③ URI로 무언가를 찾았을 때, 표준 기술(RDF, SPARQL)을 사용하여 유용한 정보를 제공하라.

④ 데이터는 반드시 XML 형식으로만 제공하라.

정답 4번

출제 포인트

링크드 데이터의 기본 원칙인 URI 식별, HTTP URI 사용, 표준 기술을 통한 정보 제공과 데이터 형식의 관계를 묻는 문제이다.

핵심 개념

링크드 데이터는 웹상의 대상을 URI로 식별하고, HTTP URI로 조회할 수 있게 하며, 조회 결과를 RDF·SPARQL 등의 표준 기술로 제공하고 다른 URI와 연결하는 데이터 공개 방식이다. 특정 직렬화 형식 하나만을 강제하는 개념이 아니다.

선지별 해설

① O — 링크드 데이터는 사물이나 개념 등 식별하려는 대상에 URI를 부여한다. URI는 웹에서 각 대상을 고유하게 가리키는 이름 역할을 한다.
② O — HTTP URI를 사용하면 해당 URI를 웹에서 조회하여 그 대상에 관한 정보에 접근할 수 있다.
③ O — URI를 조회했을 때 RDF와 SPARQL 같은 표준 기술을 이용해 유용한 정보를 제공하는 것은 링크드 데이터의 기본 원칙에 해당한다.
④ X — 틀린 표현은 '반드시 XML 형식으로만'이다. 링크드 데이터의 핵심은 URI와 RDF 기반의 의미 표현 및 연결이며, 데이터의 직렬화 형식을 XML 하나로 제한하지 않는다. 올바르게는 '데이터는 RDF 등 표준적인 방식으로 제공하되 XML 형식에만 한정되지 않는다'이다.

오답 포인트

RDF/XML이 RDF를 표현하는 방식 중 하나라는 사실을 '링크드 데이터는 XML만 사용한다'로 확대하면 ④에 속기 쉽다. RDF라는 데이터 모델과 XML이라는 직렬화 형식은 구분해야 한다.

다시 나오면 이렇게 판단

URI로 대상을 식별하는지, HTTP URI로 조회할 수 있는지, 조회 시 표준 형식의 정보를 제공하는지, 다른 URI와 연결하는지를 차례로 확인한다. '반드시', '오직', '한 형식만'처럼 개방성을 제한하는 표현은 우선 의심한다.

NoteBase

암기 포인트

링크드 데이터는 URI로 이름 붙이고, HTTP로 찾고, RDF로 설명하고, 다른 URI로 연결한다. XML 전용이 아니다.

NoteBase

문 5. 다음 Creative Commons(CC) 라이선스 유형 중, 저작물의 비영리적 이용만을 허용하며, 2차적 저작물 작성 시 원저작물과 동일한 라이선스 조건을 적용해야 하는 라이선스는?

① CC BY-NC

② CC BY-ND

③ CC BY-NC-SA

④ CC BY-SA

정답 3번

출제 포인트

Creative Commons 라이선스의 조건인 BY(저작자 표시), NC(비영리), ND(변경 금지), SA(동일조건변경허락)를 조합하여 구별하는 문제이다.

핵심 개념

BY는 저작자를 표시해야 한다는 조건이고, NC는 영리 목적으로 이용할 수 없다는 조건이다. ND는 2차적 저작물의 작성을 허용하지 않는 조건이며, SA는 2차적 저작물을 배포할 때 원저작물과 동일한 라이선스 조건을 적용하는 조건이다. 따라서 비영리 이용과 동일조건변경허락을 함께 요구하는 조합은 BY-NC-SA이다.

선지별 해설

① X — CC BY-NC에는 NC 조건은 있지만 SA 조건이 없으므로 2차적 저작물에 동일한 라이선스를 적용할 의무가 없다. 틀린 부분은 동일한 라이선스 조건까지 요구한다고 볼 수 있다는 점이며, 올바른 라이선스는 CC BY-NC-SA이다.
② X — CC BY-ND의 ND는 2차적 저작물의 작성을 허용하지 않는 조건이고, NC 조건도 없다. 틀린 부분은 비영리 이용과 동일조건 적용을 허용한다고 보는 것이며, 올바른 라이선스는 CC BY-NC-SA이다.
③ O — CC BY-NC-SA는 저작자 표시(BY), 비영리(NC), 동일조건변경허락(SA)을 모두 포함하므로 문제의 두 조건을 모두 충족한다.
④ X — CC BY-SA는 2차적 저작물에 동일한 라이선스를 적용하도록 하지만 NC 조건이 없어 영리적 이용도 허용한다. 틀린 부분은 비영리적 이용만 허용한다고 보는 것이며, 비영리 조건까지 포함하려면 CC BY-NC-SA이어야 한다.

오답 포인트

①은 NC만 보고 고르기 쉽고, ④는 SA만 보고 고르기 쉽다. 문제는 비영리와 동일조건 적용을 동시에 요구하므로 NC와 SA가 모두 포함된 조합을 찾아야 한다. 또한 ND는 동일조건으로 변경을 허용하는 조건이 아니라 변경 자체를 허용하지 않는 조건이다.

NoteBase

다시 나오면 이렇게 판단

문장의 조건을 약어로 하나씩 바꾼다. 비영리는 NC, 변경 금지는 ND, 변경을 허용하되 같은 조건을 유지하는 것은 SA이므로, 비영리와 동일조건이 함께 나오면 NC와 SA가 모두 있는지를 확인한다.

암기 포인트

NC는 비영리, ND는 변경 금지, SA는 바꿔도 같은 조건이다. 비영리이면서 동일조건이면 BY-NC-SA이다.

NoteBase

문 6. 다음 중 1974년 미국도서관정보위원회(NCLIS)에서 정보 리터러시라는 용어가 처음 사용되었던 당시의 정보격차 개념에 가장 가까운 내용은?

① 정보 활용 능력의 차이로 인해 발생하는 다양한 사회경제적 불평등 심화 현상을 포괄적으로 지칭한다.

② 개인이 정보를 효과적으로 검색하고, 비판적으로 평가하며, 윤리적으로 이용하는 능력의 차이를 주로 의미한다.

③ 1차 자료의 이용 및 새로운 정보통신기술이나 다양한 정보 도구를 활용하는 기법과 기능에 초점을 맞추어 논의되었다.

④ 정보 생산 주체와 정보 소비 주체 간의 정보 생성량 및 정보 공유 방식의 현저한 불균형 상태를 의미한다.

정답 3번

출제 포인트

1974년 정보 리터러시 개념이 처음 제시될 당시의 초기 의미와 이후 확장된 정보 리터러시·정보격차 개념을 구별하는 문제이다.

핵심 개념

초기의 정보 리터러시는 1차 자료를 이용하고 새로운 정보통신기술과 여러 정보 도구를 다루는 실용적인 기법·기능에 초점을 두었다. 이후에는 정보의 검색·평가·윤리적 이용을 아우르는 역량으로 확장되었고, 정보격차도 접근이나 활용 능력의 차이에서 비롯되는 사회경제적 불평등까지 포괄하는 의미로 넓어졌다.

선지별 해설

① X — '다양한 사회경제적 불평등 심화 현상을 포괄적으로 지칭한다'는 표현은 정보격차 개념이 넓어진 뒤의 설명이다. 올바르게는 초기 논의가 정보 자료와 새로운 정보기술·도구를 실제로 이용하는 기법과 기능에 초점을 두었다고 해야 한다.
② X — '비판적으로 평가하며, 윤리적으로 이용하는 능력'은 이후 정교화된 현대적 정보 리터러시의 포괄적 역량을 설명한다. 올바르게는 당시에는 1차 자료와 새로운 정보기술·도구의 활용 기능이 중심이었다.
③ O — 1차 자료의 이용과 새로운 정보통신기술 및 여러 정보 도구를 활용하는 기법·기능을 강조한 내용으로, 1974년 당시의 초기 개념에 가장 가깝다.
④ X — '정보 생산 주체와 정보 소비 주체 간의 정보 생성량 및 공유 방식의 불균형'은 당시 정보 리터러시가 주목한 핵심이 아니다. 올바르게는 개인이 정보 자료와 새 정보기술·도구를 활용할 수 있는 기능의 차이에 초점을 둔 설명이어야 한다.

오답 포인트

②는 오늘날 널리 쓰이는 정보 리터러시의 정의라서 그럴듯하지만, 문항은 '1974년 처음 사용되었던 당시'의 의미를 묻는다. 시대 한정 표현을 놓치면 현대적 정의를 고르게 된다.

NoteBase

다시 나오면 이렇게 판단

초기 개념을 물으면 자료·기술·도구를 실제로 다루는 기능에 주목하고, 현대적 개념을 물으면 검색·평가·윤리적 이용까지 포함하는 종합 역량에 주목한다.

암기 포인트

1974년 초기 정보 리터러시의 핵심은 '1차 자료 + 새 정보기술·도구의 활용 기능'이다.

NoteBase

문 7. 정보학에서 온톨로지란 특정 도메인에 존재하는 실체의 속성을 파악하여 개념으로 표현하고 개념 사이의 다양한 의미관계를 표현함으로써 의미 기반 추론 및 지식 모델링을 가능하게 한다. 기계가독형 온톨로지를 구축하기 위해서는 전용 언어가 개발되어야 한다. 다음 중 국제표준기구(ISO)가 제안한 온톨로지 언어로 지식정보를 의미적 연관성에 따라 구조화함으로써 대용량의 지식정보를 효율적으로 검색하고 관리해 주는 것을 목적으로 제시된 것은?

① OIL(Ontology Interchange Language)

② OWL(Ontology Web Language)

③ RDF/RDFS(Resource Description Framework/RDF Schema)

④ Topic Maps

정답 4번

출제 포인트

온톨로지 관련 언어·표준 가운데 ISO가 제안한 지식 구조화 표준인 Topic Maps를 구별하는 문제이다.

핵심 개념

Topic Maps는 주제(topic), 주제 사이의 연관(association), 주제에 관한 정보 자원(occurrence)을 이용해 지식의 의미적 관계를 구조화하는 ISO 표준이다.
OIL은 웹 온톨로지 표현 언어, OWL은 W3C의 웹 온톨로지 언어, RDF/RDFS는 W3C의 자원 기술 및 어휘 정의 기반이다. 이들을 가르는 핵심 기준은 제안·표준화 주체와 주된 표현 목적이다.

선지별 해설

① X — OIL은 웹에서 온톨로지를 표현하기 위해 개발된 언어로, 문항이 묻는 ISO의 Topic Maps가 아니다. 틀린 표현은 'ISO가 제안한 언어'에 OIL을 대응한 것이며, 올바르게는 Topic Maps이다.
② X — OWL의 정식 명칭은 Web Ontology Language이며 W3C가 표준화한 웹 온톨로지 언어이다. 틀린 표현은 OWL을 ISO가 제안한 언어로 본 것이며, 올바르게는 Topic Maps이다.
③ X — RDF는 웹 자원과 그 관계를 기술하는 W3C 기반 모델이고, RDFS는 RDF 어휘의 클래스와 속성 등을 정의하는 스키마 언어이다. 틀린 표현은 RDF/RDFS를 ISO가 제안한 해당 언어로 본 것이며, 올바르게는 Topic Maps이다.
④ O — Topic Maps는 지식정보를 주제와 주제 간 연관관계 중심으로 구조화하여 검색·관리할 수 있게 하는 ISO 표준이므로 문항의 설명에 해당한다.

오답 포인트

OIL, OWL, RDF/RDFS도 모두 의미 기반의 지식 표현과 관련되어 있어 기능만 보면 혼동하기 쉽다. 이 문항의 결정적 단서는 'ISO가 제안'했다는 표현과 주제 사이의 의미적 연관성에 따른 구조화이다.

NoteBase

다시 나오면 이렇게 판단

표준화 주체와 핵심 단위를 함께 확인한다. ISO와 주제·연관·정보 자원이 제시되면 Topic Maps, W3C와 웹 온톨로지가 제시되면 OWL, 자원 기술과 스키마가 제시되면 RDF/RDFS로 판단한다.

암기 포인트

ISO의 지식 지도는 Topic Maps, W3C의 웹 온톨로지는 OWL이다.

NoteBase

문 8. 어떤 학술 데이터베이스에서 특정 주제 분야의 논문 출판 경향을 Lotka의 법칙을 이용하여 분석하였다. 이 분야에서 1편의 논문을 발표한 저자가 총 900명으로 집계되었을 때, Lotka의 법칙에 따른다면 4편의 논문을 발표한 저자의 수는 약 몇 명으로 예상할 수 있는가? (단, 계산 결과는 정수로 반올림한다)

① 약 28명

② 약 56명

③ 약 113명

④ 약 225명

정답 2번

출제 포인트

Lotka의 법칙을 적용하여 논문 수가 늘어날 때 해당 논문 수를 발표한 저자 수를 계산하는 문제이다.

조건 정리

주어진 값: 1편을 발표한 저자는 900명이다.
구해야 할 값: 4편을 발표한 저자의 수이다.
주의할 조건: 계산 결과를 정수로 반올림한다.

사용할 공식

Lotka의 법칙에서 n편을 발표한 저자의 수를 A_n, 1편을 발표한 저자의 수를 A_1이라 하면 다음과 같다.
$A_n = \frac{A_1}{n^2}$
저자 수가 발표 논문 수의 제곱에 반비례하므로 이 공식을 사용한다.

계산 과정

4편을 발표한 저자의 수는 다음과 같다.
$A_4 = \frac{900}{4^2} = \frac{900}{16} = 56.25$
56.25를 정수로 반올림하면 약 56명이다.

선지 판단

① X — ‘약 28명’은 계산값 56.25를 반올림한 수가 아니다. 올바르게는 ‘약 56명’이다.
② O — 56.25를 정수로 반올림하면 약 56명이므로 옳다.
③ X — ‘약 113명’은 900을 4^2으로 나눈 값이 아니다. 올바르게는 ‘약 56명’이다.
④ X — ‘약 225명’은 900을 4로만 나눈 값으로, 논문 수의 제곱에 반비례한다는 조건을 반영하지 못한 수이다. 올바르게는 ‘약 56명’이다.

NoteBase

오답 포인트

Lotka의 법칙에서 저자 수는 논문 수에 단순 반비례하는 것이 아니라 논문 수의 제곱에 반비례한다. ④의 225명은 900을 4로만 나눈 값이므로 제곱을 빠뜨린 함정이다.

다시 나오면 이렇게 판단

1편 저자 수를 기준으로 n편 저자 수를 묻는 경우 기준 저자 수를 n^2으로 나눈 뒤, 문제에서 요구한 방식으로 반올림한다.

암기 포인트

Lotka의 법칙은 '논문 수 n배, 저자 수 $1/n^2$배'이다.

NoteBase

문 9. 다음 중 정보 검색 시스템의 성능을 평가하는 가장 기본적인 두 가지 척도는?

① 속도와 안정성

② 정확률과 재현율

③ 다양성과 최신성

④ 이용 편의성과 디자인

정답 2번

출제 포인트

정보 검색 시스템의 검색 효과를 평가하는 기본 척도인 정확률과 재현율의 의미를 묻는 문제이다.

핵심 개념

정확률: 검색된 문서 가운데 실제로 적합한 문서가 차지하는 비율이다. 검색 결과의 정확성을 나타낸다.
재현율: 전체 적합 문서 가운데 검색 시스템이 찾아낸 문서가 차지하는 비율이다. 적합 문서를 얼마나 빠짐없이 검색했는지를 나타낸다.
정확률은 '검색 결과 중 적합 문서', 재현율은 '전체 적합 문서 중 검색된 문서'를 기준으로 한다.

선지별 해설

① X — '속도와 안정성'은 응답 시간이나 시스템 운영 상태와 관련된 항목이지만, 검색 결과의 적합성을 평가하는 가장 기본적인 두 척도는 아니다. 올바르게는 '정확률과 재현율'이다.
② O — 정확률은 검색된 문서 중 적합 문서의 비율이고, 재현율은 전체 적합 문서 중 검색된 문서의 비율이므로 정보 검색 성능을 평가하는 기본적인 한 쌍이다.
③ X — '다양성과 최신성'은 검색 결과의 품질을 살필 때 고려할 수 있는 속성이지만, 정보 검색 성능의 가장 기본적인 두 척도는 아니다. 올바르게는 '정확률과 재현율'이다.
④ X — '이용 편의성과 디자인'은 사용자 인터페이스와 사용성에 관한 요소이므로 검색 결과가 얼마나 적합하고 빠짐없는지를 직접 측정하지 않는다. 올바르게는 '정확률과 재현율'이다.

오답 포인트

속도·안정성, 다양성·최신성, 이용 편의성·디자인도 시스템 평가 요소처럼 보이지만, 이 문항은 검색 결과의 적합성을 측정하는 고전적인 기본 척도 한 쌍을 묻는다. 정확률과 재현율의 분모가 서로 다르다는 점도 함께 구분해야 한다.

다시 나오면 이렇게 판단

정보 검색의 기본 성능 척도를 물으면 검색 결과의 정확성을 보는 정확률과 적합 문서의 누락 정도를 보는 재현율을 찾는다. '검색된 것 중 맞는 것'은 정확률, '맞는 것 전체 중 찾아낸 것'은 재현율로 판단한다.

NoteBase

암기 포인트

정확률은 찾아낸 것의 정확성, 재현율은 찾아내야 할 것의 회수성이다.

NoteBase

문 10. 다음 중 정보 검색에서 용어 가중치를 부여할 때 사용되는 역문헌빈도의 역할과 그 계산 방식의 근본적인 이유로 가장 옳은 설명은?

① 특정 문서 내에서 특정 용어가 자주 등장할수록 그 용어의 중요도가 높다고 판단하여 해당 용어에 높은 가중치를 부여한다.

② 문서의 길이에 따라 용어의 가중치가 과도하게 평가되는 것을 방지하기 위해, 모든 문서의 길이를 동일하게 정규화한다.

③ 전체 문서 집합에서 특정 용어가 나타난 문헌의 수에 역수를 취하고 로그를 적용함으로써, 여러 문서에 걸쳐 보편적으로 등장하는 용어의 변별력을 낮추고 특정 주제를 잘 나타내는 용어의 가중치를 상대적으로 높인다.

④ 사용자의 질의에 명시적으로 포함된 용어에 대해서만 가중치를 부여하고, 질의에 없는 용어는 가중치를 0으로 처리한다.

정답 3번

출제 포인트

역문헌빈도(IDF)가 문서 집합 전체에서 용어의 희소성을 반영하는 이유와 용어빈도(TF), 문서 길이 정규화, 질의 용어 처리의 차이를 묻는 문제이다.

핵심 개념

역문헌빈도(IDF)는 전체 문서 수를 N, 용어 t가 등장한 문서 수를 $df(t)$라고 할 때 보통 $IDF(t) = \log\left(\frac{N}{df(t)}\right)$의 형태로 계산한다. 많은 문서에 흔히 등장하는 용어는 $df(t)$가 커져 IDF가 낮아지고, 소수 문서에만 등장하는 용어는 IDF가 높아져 문서를 구별하는 힘이 크게 평가된다. 반면 용어빈도(TF)는 한 문서 안에서 용어가 얼마나 자주 등장하는지를 나타낸다.

선지별 해설

① X — '특정 문서 내에서 특정 용어가 자주 등장할수록' 중요도를 높이는 것은 역문헌빈도가 아니라 용어빈도(TF)의 설명이다. 올바르게는 IDF는 그 용어가 등장한 문서 수가 적을수록 높은 가중치를 부여한다.
② X — '모든 문서의 길이를 동일하게 정규화한다'는 것은 IDF의 계산 방식이 아니다. 문서 길이 정규화는 긴 문서가 단어 출현 횟수 때문에 유리해지는 현상을 조정하는 별도 처리이며, IDF는 전체 문서 수와 해당 용어가 등장한 문서 수의 비율을 이용한다.
③ O — $df(t)$에 반비례하는 비율에 로그를 적용하면 여러 문서에 흔히 나타나는 용어의 가중치는 낮아지고, 특정 문서군의 주제를 잘 구별하는 희소한 용어의 가중치는 상대적으로 높아진다. 이것이 IDF의 역할과 계산 이유이다.
④ X — '질의에 명시적으로 포함된 용어에 대해서만' 및 '질의에 없는 용어는 가중치를 0으로 처리한다'는 것은 IDF의 정의가 아니다. 올바르게는 IDF는 질의 포함 여부가 아니라 전체 문서 집합에서 각 용어가 몇 개 문서에 나타나는지를 기준으로 정해진다.

NoteBase

오답 포인트

①의 문서 내부 출현 빈도는 TF, ②의 문서 길이 조정은 정규화인데 이를 IDF와 혼동하기 쉽다. ③에서는 단순히 용어가 많이 나온다는 점이 아니라 '여러 문서에 걸쳐 흔한가'를 기준으로 변별력을 낮춘다는 표현이 핵심이다.

다시 나오면 이렇게 판단

한 문서 안의 반복 횟수를 묻으면 TF, 전체 문서 중 그 용어가 나온 문서의 수를 묻으면 IDF로 판단한다.
IDF에서는 문서빈도가 클수록 가중치가 작아지고 문서빈도가 작을수록 가중치가 커진다.

암기 포인트

TF는 문서 안에서 자주, IDF는 문서 집합에서 드물수록 높다.

NoteBase

문 11. 전자정보원의 특징에 관한 다음 설명 중 옳은 것을 모두 고른 것은?

ㄱ. 정보기술, 인터넷에 의존하여 제작과 편집이 신속하다.
ㄴ. 유사 주제의 전자자원이 많은 경우 적절한 것을 선택하는 데 어려움이 있다.
ㄷ. 색인 검색, 초록과 원문 자체의 검색 지원 및 원문 접근이 쉽다.
ㄹ. 데이터베이스 간에 타이틀이 중복되지 않는다.

① ㄱ
② ㄱ, ㄴ
③ ㄱ, ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ, ㄹ

정답 3번

출제 포인트

전자정보원의 제작·검색·접근상 장점과 자원 선택·중복 문제를 구별하는 문항이다.

핵심 개념

전자정보원은 디지털 형태로 제작·유통되어 신속하게 편집할 수 있고, 색인·초록·원문 검색과 원문 접근이 편리한 정보원이다. 다만 유사한 자원이 많아 적절한 자원을 선택하기 어렵고, 여러 데이터베이스에 같은 타이틀이 중복 수록될 수 있다. 따라서 ㄱ, ㄴ, ㄷ은 옳고 ㄹ은 틀리다.

선지별 해설

① X — ‘ㄱ’만 고른 것이 틀렸다. ㄱ뿐 아니라 ㄴ과 ㄷ도 전자정보원의 특징이므로 올바른 조합은 ‘ㄱ, ㄴ, ㄷ’이다.
② X — ‘ㄱ, ㄴ’에서 ㄷ을 제외한 것이 틀렸다. 전자정보원은 색인·초록·원문 검색을 지원하고 원문 접근도 쉬우므로 올바른 조합은 ‘ㄱ, ㄴ, ㄷ’이다.
③ O — ㄱ의 신속한 제작·편집, ㄴ의 자원 선택 곤란, ㄷ의 편리한 검색·원문 접근은 모두 전자정보원의 특징이다.
④ X — ㄹ을 포함한 것이 틀렸다. 데이터베이스 간에는 같은 타이틀이 중복 수록될 수 있으므로 ‘타이틀이 중복되지 않는다’가 아니라 ‘타이틀이 중복될 수 있다’가 옳다. 따라서 올바른 조합은 ‘ㄱ, ㄴ, ㄷ’이다.

오답 포인트

ㄹ의 ‘중복되지 않는다’라는 단정이 함정이다. 전자정보원은 검색과 접근이 편리하지만, 데이터베이스마다 수록 범위가 겹쳐 동일 타이틀이 중복될 수 있다.

NoteBase

다시 나오면 이렇게 판단

신속한 갱신·다양한 검색·쉬운 접근은 장점으로 판단하고, 유사 자원의 선택 곤란과 데이터베이스 간 타이틀 중복은 관리상 문제로 판단한다. '중복이 없다'처럼 문제점을 전면 부정하는 표현을 점검한다.

암기 포인트

전자정보원은 제작·검색·접근은 빠르고 편리하지만, 선택은 어렵고 타이틀은 중복될 수 있다.

NoteBase

문 12. 다음 중 오픈액세스(OA)의 실현을 위한 두 가지 핵심 전략으로 부다페스트 선언(BOAI)에서 제시된 것은?

① 골드 OA 학술지 출판과 다이아몬드 OA 리포지터리 구축을 통해 모든 연구 결과물을 즉시 공개하는 전략이다.

② 셀프 아카이빙을 통해 연구자가 자신의 논문을 리포지터리에 공개하고, OA 학술지를 통해 논문을 출판하는 전략이다.

③ 연구데이터의 완전한 공개를 의무화하고, 모든 학술 출판물에 CC0(Creative Commons Zero) 라이선스를 적용하는 전략이다.

④ 정부 주도의 중앙 집중형 학술 정보 포털을 구축하고, 모든 연구자에게 논문처리비용(APC) 면제 혜택을 제공하는 전략이다.

정답 2번

출제 포인트

부다페스트 오픈액세스 선언(BOAI)이 제시한 두 핵심 전략인 셀프 아카이빙과 오픈액세스 학술지 출판을 구별하는 문제이다.

핵심 개념

셀프 아카이빙은 연구자가 자신의 논문을 공개 리포지터리에 저장하여 누구나 접근할 수 있게 하는 전략이다. 흔히 그린 OA와 연결된다.
OA 학술지 출판은 독자가 무료로 이용할 수 있는 오픈액세스 학술지에 논문을 출판하는 전략이다. 흔히 골드 OA와 연결된다.
BOAI의 핵심 구도는 '리포지터리에 직접 공개'와 'OA 학술지에 출판'이다.

선지별 해설

① X — '골드 OA 학술지 출판과 다이아몬드 OA 리포지터리 구축'이라는 구분이 틀렸다. BOAI가 제시한 두 전략은 셀프 아카이빙과 OA 학술지 출판이며, 올바르게는 '연구자가 논문을 리포지터리에 공개하고 OA 학술지에 출판한다'이다.
② O — 연구자가 자신의 논문을 리포지터리에 공개하는 셀프 아카이빙과 OA 학술지를 통한 출판을 정확히 제시한다.
③ X — '연구데이터의 완전한 공개 의무화'와 '모든 학술 출판물에 CC0 적용'이 틀렸다. BOAI의 두 핵심 전략은 논문의 셀프 아카이빙과 OA 학술지 출판이며, 모든 연구데이터 공개나 일률적인 CC0 적용을 두 전략으로 제시한 것이 아니다.
④ X — '정부 주도의 중앙 집중형 포털 구축'과 '모든 연구자에 대한 APC 면제'가 틀렸다. 올바르게는 분산된 리포지터리를 활용한 셀프 아카이빙과 OA 학술지를 통한 출판이 핵심 전략이다.

NoteBase

오답 포인트

①은 골드 OA라는 익숙한 용어에 '다이아몬드 OA 리포지터리'를 섞어 정답처럼 보이게 한다. 다이아몬드 OA는 일반적으로 저자에게도 논문처리비용을 받지 않는 OA 출판 모델을 가리키며, BOAI의 셀프 아카이빙 전략을 대신하는 명칭이 아니다.

다시 나오면 이렇게 판단

BOAI의 두 전략을 묻는 경우 공개 장소를 기준으로 판단한다. 연구자가 리포지터리에 올리면 셀프 아카이빙이고, OA 학술지 자체에 출판하면 OA 학술지 전략이다.

암기 포인트

BOAI는 '직접 저장(셀프 아카이빙) + 공개 학술지(OA 학술지)'의 두 길이다.

NoteBase

문 13. 저널의 피인용지수(Impact Factor, IF)에 관한 다음 설명 중 가장 옳은 것은?

① 개별 논문의 질적 수준을 직접적으로 평가하는 가장 신뢰도 높은 지표이다.

② 해당 저널에 지난 5년간 게재된 논문들이 특정 연도에 얼마나 평균적으로 인용되었는지를 나타낸다.

③ 특정 연도(예: 2024년)의 IF는 해당 저널에 2022년과 2023년에 게재된 논문들이 2024년 한 해 동안 인용된 총 횟수를, 2022년과 2023년 두 해 동안 해당 저널에 게재된 총 인용 가능 논문 수로 나눈 값이다.

④ IF는 주로 논문 저자의 학문적 영향력을 평가하기 위해 개발된 지표이다.

정답 3번

출제 포인트

저널 피인용지수(Impact Factor, IF)의 평가 대상과 산출 기간을 정확히 구분하는 문제이다.

핵심 개념

IF는 특정 저널에 실린 논문들이 얼마나 자주 인용되는지를 저널 단위의 평균으로 나타내는 지표이다. 특정 연도 Y의 IF는 $Y-1$년과 $Y-2$년에 그 저널에 게재된 논문이 Y년에 받은 인용 횟수를 같은 두 해의 인용 가능 논문 수로 나누어 산출한다. 따라서 2024년 IF의 산출 대상 게재 연도는 2022년과 2023년이다.

선지별 해설

① X — 틀린 표현은 '개별 논문의 질적 수준을 직접적으로 평가하는 가장 신뢰도 높은 지표'이다. IF는 저널 단위의 평균 인용 영향력을 나타내며, 개별 논문의 질을 직접 평가하는 지표가 아니다.
② X — 틀린 표현은 '지난 5년간 게재된 논문들'이다. 일반적인 IF는 특정 연도의 직전 2개 연도에 게재된 인용 가능 논문을 기준으로 산출한다. 5년간의 논문을 대상으로 하는 값은 별도의 5-year Impact Factor이다.
③ O — 2024년에 2022년과 2023년 게재 논문이 받은 인용 횟수를 분자로 하고, 그 두 해에 게재된 인용 가능 논문 수를 분모로 한다는 설명은 IF의 산출 방식과 일치한다.
④ X — 틀린 표현은 '논문 저자의 학문적 영향력을 평가'이다. IF의 주된 평가 단위는 연구자 개인이 아니라 저널이다. 연구자나 개별 논문의 영향력을 IF와 동일시해서는 안 된다.

오답 포인트

IF가 인용을 이용한다는 이유로 이를 개별 논문이나 저자의 질·영향력 지표로 오해하기 쉽다. 또한 일반 IF의 2년 산출 창과 별도의 5년 IF를 혼동하는 것이 핵심 함정이다.

다시 나오면 이렇게 판단

평가 단위는 저널인지, 게재 기간은 직전 2개 연도인지, 인용이 집계되는 기간은 산출 연도 1년인지 차례로 확인한다.

NoteBase

암기 포인트

IF는 '저널 단위·직전 2년 게재분·해당 연도 인용'이다.

NoteBase

문 14. 다음 중 정보검색의 발전에 기여가 큰 연구자 가운데 정보검색(information retrieval), 디스크립터(descriptor)라는 용어를 처음으로 만들고 '정보를 검색하고 소유하는 것이 그렇지 않을 때보다 더 고통스럽고 귀찮을 때 정보검색 시스템은 사용되지 않는다'는 법칙을 제시한 사람은 누구인가?

① B.C. Brookes

② Eugene Garfield

③ Calvin N. Mooers

④ Gerard Salton

정답 3번

출제 포인트

정보검색·디스크립터라는 용어와 '정보검색 시스템이 번거로우면 이용되지 않는다'는 무어스의 법칙을 제시한 연구자를 연결하는 문제이다.

핵심 정리

B.C. Brookes: 정보과학의 기본방정식을 통해 지식구조가 정보의 수용으로 변화하는 과정을 설명하였다.
Eugene Garfield: 인용색인과 과학인용색인의 발전에 기여하여 문헌 간 인용 관계를 정보검색에 활용하였다.
Calvin N. Mooers: information retrieval과 descriptor라는 용어를 만들었으며, 정보 획득의 수고가 이익보다 크면 정보검색 시스템이 이용되지 않는다는 무어스의 법칙을 제시하였다.
Gerard Salton: SMART 정보검색 시스템과 벡터공간모형을 발전시켜 자동 정보검색의 토대를 마련하였다.

선지별 해설

① X — B.C. Brookes는 정보과학의 기본방정식으로 알려진 연구자이다. 틀린 부분은 제시된 용어와 법칙을 Brookes에게 연결한 것이며, 올바르게는 Calvin N. Mooers와 연결된다.
② X — Eugene Garfield는 인용색인과 과학인용색인의 발전에 기여한 연구자이다. 틀린 부분은 제시된 용어와 법칙을 Garfield에게 연결한 것이며, 올바르게는 Calvin N. Mooers와 연결된다.
③ O — Calvin N. Mooers는 information retrieval과 descriptor라는 용어를 만들고, 정보검색 시스템의 이용 여부를 이용자가 느끼는 수고와 불편의 관점에서 설명한 무어스의 법칙을 제시하였다.
④ X — Gerard Salton은 SMART 시스템과 벡터공간모형 등 자동 정보검색 연구로 알려진 인물이다. 틀린 부분은 제시된 용어와 법칙을 Salton에게 연결한 것이며, 올바르게는 Calvin N. Mooers와 연결된다.

출제자가 노린 함정

네 사람 모두 정보학과 정보검색 발전에 기여했기 때문에 업적을 서로 바꾸어 기억하기 쉽다. 특히 ③의 '용어 창안'과 '검색의 번거로움'을 하나의 묶음으로 보지 않으면 인용색인의 Garfield나 자동검색의 Salton을 고르기 쉽다.

NoteBase

다시 나오면 이렇게 판단

인물명과 대표 업적을 짝으로 구분한다. ‘information retrieval·descriptor·검색이 귀찮으면 사용하지 않음’은 Mooers, ‘인용색인’은 Garfield, ‘SMART·벡터공간모형’은 Salton, ‘정보과학의 기본방정식’은 Brookes로 판단한다.

암기 포인트

Mooers = 정보검색·디스크립터 용어 + 무어스의 법칙

NoteBase

문 15. 다음 중 의학, 약학 분야의 대표적인 서지 데이터베이스로, 특히 의약품 관련 정보에 대해 포괄적으로 제공하는 정보원은?

① MEDLINE
② Emerald
③ PubMed Central
④ EMBASE

정답 4번

출제 포인트

이 문제는 의학·약학 분야 정보원의 성격을 구별하고, 특히 의약품 정보 수록에 강한 서지 데이터베이스를 찾는 문제이다.

핵심 개념

EMBASE: 의학·생명과학 분야의 서지 데이터베이스로, 약리학과 의약품 관련 문헌을 폭넓게 수록한다.
키워드는 '의약품·약리학'이다.
MEDLINE: 미국 국립의학도서관이 구축한 대표적인 생의학 서지 데이터베이스이다.
PubMed Central: 생의학·생명과학 논문의 원문을 무료로 제공하는 디지털 아카이브이다.
Emerald: 경영학·사회과학 분야의 학술지와 전자책 등을 제공하는 학술정보 플랫폼이다.

선지별 해설

① X — MEDLINE은 대표적인 생의학 서지 데이터베이스이지만, 문제에서 강조한 '특히 의약품 관련 정보를 포괄적으로 제공'하는 정보원은 아니다. 올바르게는 EMBASE이다.
② X — Emerald는 주로 경영학·사회과학 분야의 학술정보를 제공하므로 의학·약학 분야의 대표적인 의약품 중심 서지 데이터베이스가 아니다. 올바르게는 EMBASE이다.
③ X — PubMed Central은 생의학 논문의 무료 원문을 보존·제공하는 아카이브이므로, 의약품 정보를 포괄적으로 색인하는 서지 데이터베이스라는 설명과 맞지 않는다. 올바르게는 EMBASE이다.
④ O — EMBASE는 의학·생명과학 문헌을 수록하면서 특히 약리학과 의약품 관련 정보에 강한 대표적인 서지 데이터베이스이다.

오답 포인트

MEDLINE과 PubMed Central도 모두 생의학 분야 정보원이어서 혼동하기 쉽다. 그러나 '의약품 관련 정보에 특히 포괄적'이라는 표현이 EMBASE를 가리키는 결정적 단서이다. 또한 PubMed Central은 서지 색인보다 무료 원문 아카이브라는 성격이 핵심이다.

NoteBase

다시 나오면 이렇게 판단

정보원의 주제 분야뿐 아니라 제공 형태와 강점까지 함께 구별한다. ‘의약품·약리학에 강한 서지 데이터베이스’이면 EMBASE, ‘대표적 생의학 서지 데이터베이스’이면 MEDLINE, ‘무료 원문 아카이브’이면 PubMed Central로 판단한다.

암기 포인트

EMBASE는 의약품·약리학, MEDLINE은 생의학 서지, PubMed Central은 무료 원문이다.

NoteBase

문 16. 이용자가 검색창에 '블로그'라고만 입력했을 때, 시스템이 이용자가 '블로그의 정의'를 원하는지, '인기 있는 블로그 목록'을 원하는지, 아니면 '블로그 개설 방법'을 원하는지 파악하려 한다면, 이는 자연언어 처리의 어떤 단계를 응용한 것이라고 볼 수 있는가?

① 형태소 분석(Morphological analysis)

② 구문 분석(Syntactic analysis)

③ 의미 분석(Semantic analysis)

④ 화용 분석(Pragmatic analysis)

정답 4번

출제 포인트

이 문제는 자연언어 처리 단계 중 문맥과 이용 상황을 바탕으로 발화자의 실제 의도를 파악하는 단계를 구별하는 문제이다.

핵심 개념

형태소 분석: 문장을 의미를 가진 최소 단위인 형태소로 나누고 각 형태소의 품사나 형태를 파악한다.
구문 분석: 문장 성분 사이의 문법적 관계와 문장 구조를 파악한다.
의미 분석: 단어나 문장이 나타내는 문자적 의미를 파악하고 의미 관계나 중의성을 분석한다.
화용 분석: 문맥, 상황, 화자의 목적을 고려하여 실제로 전달하려는 의도를 파악한다.
이 문항에서는 동일한 검색어 '블로그'가 정의 검색, 인기 목록 검색, 개설 방법 검색 중 무엇을 뜻하는지 이용자의 의도로 해석하므로 화용 분석에 해당한다.

선지별 해설

① X — 형태소 분석은 '블로그'라는 입력을 형태소로 나누거나 품사를 판별하는 단계이다. 이용자가 정의·목록·개설 방법 중 무엇을 원하는지 추론하는 단계가 아니므로, 올바르게는 화용 분석이다.
② X — 구문 분석은 문장 안의 문법적 구조와 성분 간 관계를 파악하는 단계이다. 한 단어로 제시된 검색어에서 이용 목적을 추론하는 단계가 아니므로, 올바르게는 화용 분석이다.
③ X — 의미 분석은 단어나 문장의 문자적 의미와 의미 관계를 파악하는 단계이다. '블로그'의 의미 자체가 아니라 그 말을 사용한 이용자의 구체적 검색 의도를 판단하므로, 올바르게는 화용 분석이다.
④ O — 화용 분석은 표현의 표면적 의미를 넘어 문맥과 사용 상황을 바탕으로 이용자의 실제 의도를 파악한다. '블로그'라는 동일한 입력에서 정의·인기 목록·개설 방법 중 원하는 바를 구별하는 것이 이에 해당한다.

오답 포인트

'블로그'라는 단어의 의미를 다룬다는 이유로 ③ 의미 분석을 고르기 쉽다. 그러나 문제의 초점은 단어의 뜻이 아니라 이용자가 그 단어로 무엇을 요구했는지에 있으므로 화용 분석이다.

NoteBase

다시 나오면 이렇게 판단

형태와 품사는 형태소 분석, 문법 구조는 구문 분석, 문자적 뜻은 의미 분석, 문맥 속 의도와 목적은 화용 분석으로 구별한다.

암기 포인트

형태소는 단어의 조각, 구문은 문장의 구조, 의미는 표현의 뜻, 화용은 상황 속 의도이다.

문 17. Shannon의 정보이론에 관한 다음 설명 중 가장 옳지 않은 것은?

① 정보의 의미 중심 처리에 기반이 되는 이론이다.

② 정보학, 커뮤니케이션, 경영 및 경제학 등에 적용되어 불확실성의 감소에 필요한 도구나 사실을 정보라고 파악한다.

③ 정보는 신호의 전달 과정을 확률적으로 적용할 수 있기에 정보에 대한 양적 측정이 가능하다고 하였다.

④ 정보는 엔트로피를 감소시키는 도구로, 엔트로피는 열역학 제2법칙의 개념이다.

정답 1번

출제 포인트

Shannon의 정보이론이 정보의 의미가 아니라 신호의 확률과 불확실성을 정량적으로 다루는 이론임을 묻는 문제이다.

핵심 개념

Shannon의 정보이론은 통신 과정에서 메시지의 의미와 무관하게 신호가 지닌 불확실성과 정보량을 확률적으로 측정하는 이론이다. 정보량이 커질수록 수신 전의 불확실성이 더 크게 줄어들며, 엔트로피는 이러한 평균 불확실성을 나타낸다.

선지별 해설

① X - 틀린 표현은 '정보의 의미 중심 처리'이다. Shannon의 정보이론은 정보의 의미나 가치가 아니라 신호의 선택 가능성과 발생 확률을 중심으로 정보량을 정량화한다. 올바르게는 '정보의 확률적·양적 처리에 기반이 되는 이론'이라고 해야 한다.
② O - 정보는 수신자의 불확실성을 줄이는 것으로 이해할 수 있으며, 이 관점은 정보학과 커뮤니케이션을 비롯한 여러 분야에 적용된다.
③ O - 신호의 발생 확률을 이용하면 정보량과 평균 정보량인 엔트로피를 수치로 계산할 수 있으므로 정보의 양적 측정이 가능하다.
④ O - 정보는 불확실성, 즉 정보 엔트로피를 감소시키는 역할을 한다. 엔트로피라는 용어와 수학적 착상은 열역학에서 유래했으며, 열역학 제2법칙과 관련된 개념이다.

오답 포인트

일상에서 정보는 의미 있는 내용으로 이해되기 때문에 ①의 '의미 중심'이 자연스럽게 보일 수 있다. 그러나 Shannon 이론은 의미를 배제하고 메시지의 발생 확률과 불확실성만을 측정한다는 점이 핵심 함정이다.

NoteBase

다시 나오면 이렇게 판단

Shannon 정보이론에 관한 선지는 '의미·가치·유용성'보다 '확률·불확실성·정보량·엔트로피·신호 전달'을 중심으로 판단한다. 의미 자체를 정량화한다고 하면 틀린 설명이다.

암기 포인트

Shannon 정보이론은 의미가 아니라 확률, 정보량은 불확실성의 감소이다.

문 18. Dublin Core 메타데이터의 특징에 관한 다음 설명 중 가장 옳지 않은 것은?

① 단순성: 15개의 핵심 요소로 구성되어 이해하고 적용하기 쉽다.

② 확장성: 한정어(Qualifier)를 사용하여 요소의 의미를 구체화하거나 확장할 수 있다.

③ 상호운용성: 다양한 시스템과 자원 유형에 적용 가능하여 정보 공유 및 교환에 유리하다.

④ 범용성: MARC와 완벽하게 호환성을 가지므로 범용적으로 사용될 수 있다.

정답 4번

출제 포인트

Dublin Core 메타데이터의 단순성, 확장성, 상호운용성을 이해하고 MARC와의 관계를 구별하는 문제이다.

핵심 개념

Dublin Core는 다양한 정보자원을 간단하고 일관되게 기술하기 위한 범용 메타데이터 요소 집합이다. 15개 핵심 요소를 바탕으로 하며, 한정어를 통해 요소의 의미나 값의 표현 방식을 더 구체화할 수 있다. 서로 다른 시스템 사이의 메타데이터 교환에는 유리하지만, 구조와 상세 수준이 다른 MARC와 완전한 일대일 호환 관계인 것은 아니다.

선지별 해설

① O — Dublin Core는 Title, Creator, Subject 등을 포함한 15개의 핵심 요소로 구성되어 구조가 단순하고 이해·적용이 쉽다.
② O — 한정어(Qualifier)는 기본 요소의 의미를 더 세밀하게 하거나 값의 표현 방식을 지정하므로 Dublin Core의 확장성을 높인다.
③ O — 특정 시스템이나 특정 자원 유형에만 한정되지 않고 여러 분야의 정보자원에 적용할 수 있어 메타데이터의 공유와 교환에 유리하다.
④ X — 틀린 표현은 ‘MARC와 완벽하게 호환성’이다. Dublin Core와 MARC 사이에는 요소 대응이나 변환이 가능하지만, 두 형식은 구조와 기술의 상세 수준이 달라 완전한 일대일 호환이 보장되지 않는다. 올바르게는 ‘Dublin Core는 단순하고 범분야적으로 적용할 수 있어 범용성이 높으며, MARC와는 매핑하여 연계할 수 있다’이다.

오답 포인트

④는 ‘범용적으로 사용된다’는 결론 자체보다 그 근거인 ‘MARC와 완벽하게 호환된다’가 함정이다. 메타데이터 형식 사이에 매핑이나 변환이 가능하다는 사실을 완전한 호환성과 동일하게 보면 안 된다.

NoteBase

다시 나오면 이렇게 판단

Dublin Core의 특징은 '15개 핵심 요소에 따른 단순성, 한정어에 따른 확장성, 여러 시스템·자원에 적용되는 상호운용성'으로 판단한다. 다른 메타데이터 형식과의 관계에서 '완벽한 호환'이나 '완전한 일대일 대응'처럼 절대적인 표현이 나오면 형식 간 구조와 상세 수준의 차이를 확인한다.

암기 포인트

Dublin Core는 15개 요소로 단순하고, 한정어로 확장하며, MARC와는 매핑할 수 있지만 완전 호환되지는 않는다.

NoteBase

문 19. 다음은 어느 연구자가 출판한 논문 목록과 각 논문의 피인용 횟수이다. 이 연구자의 h-index는 얼마이며, 이는 무엇을 의미하는가?

논문 명	논문 순번(피인용이 많은 순)	피인용 횟수
A	1	45
B	2	30
C	3	18
D	4	12
E	5	9
F	6	7
G	7	5
H	8	3

① h-index는 5이며, 이는 이 연구자가 발표한 논문 중 최소 5편의 논문이 각각 평균 9회 이상 인용되었음을 의미한다.

② h-index는 6이며, 이는 이 연구자가 발표한 논문 중 최소 6편의 논문이 각각 6회 이상 인용되었음을 의미하며, 6번째 논문의 피인용 횟수가 7회이므로 h-index는 6이 된다.

③ h-index는 7이며, 이는 이 연구자가 발표한 논문 중 7번째 논문의 피인용 횟수인 5회와 순위 7을 비교하여, 피인용 횟수가 순위보다 작아지는 직전의 순위가 h-index가 되므로 7이 된다.

④ h-index는 9이며, 이는 이 연구자가 발표한 논문 중 상위 5편의 논문이 각각 9회 이상 인용되었으므로, 5번째 논문의 피인용 횟수인 9가 h-index가 된다.

정답 2번

출제 포인트

피인용 횟수를 내림차순으로 놓았을 때, 순위와 피인용 횟수를 비교하여 h-index의 정의를 적용하는 문제이다.

조건 정리

논문의 순위별 피인용 횟수는 45, 30, 18, 12, 9, 7, 5, 3회이다. h-index는 피인용 횟수가 각각 h회 이상인 논문을 최소 h편 가진 경우의 가장 큰 h이다.

NoteBase

사용할 공식

$h = \max\{r \mid c_r \geq r\}$
여기서 r은 피인용 횟수의 내림차순 순위이고, c_r은 r번째 논문의 피인용 횟수이다. 순위 r과 해당 논문의 피인용 횟수 c_r를 직접 비교해야 하므로 이 기준을 사용한다.

계산 과정

1. 6번째 논문의 피인용 횟수는 7회이므로 c_6=7≥6이다. 따라서 최소 6편이 각각 6회 이상 인용되었다.
2. 7번째 논문의 피인용 횟수는 5회이므로 $c_7 = 5 < 7$이다. 따라서 최소 7편이 각각 7회 이상 인용되었다고 할 수 없다.
3. 기준을 만족하는 가장 큰 순위는 6이므로 h-index는 6이다.

선지 판단

① X — h-index는 5가 아니라 6이다. 또한 h-index는 평균 피인용 횟수를 뜻하지 않는다. 틀린 표현 '최소 5편이 각각 평균 9회 이상 인용'은 '최소 6편이 각각 6회 이상 인용'으로 고쳐야 한다.
② O — 6번째 논문은 7회 인용되어 6회 이상이라는 기준을 충족하고, 7번째 논문은 5회로 7회 이상이라는 기준을 충족하지 못한다. 따라서 h-index는 6이며 그 의미도 정확하다.
③ X — 7번째 논문의 피인용 횟수 5회는 순위 7보다 작으므로 7은 기준을 충족하지 못한다. 틀린 표현 'h-index는 7'은 '피인용 횟수가 순위보다 작아지는 최초 순위 7의 직전 순위인 6이 h-index이다'로 고쳐야 한다.
④ X — 5번째 논문의 피인용 횟수 9는 한 논문의 인용 횟수일 뿐 h-index가 아니다. 틀린 표현 'h-index는 9'는 '6편이 각각 6회 이상 인용되었고 7편이 각각 7회 이상 인용되지는 않았으므로 h-index는 6'으로 고쳐야 한다.

오답 포인트

h-index를 평균 피인용 횟수나 특정 순위 논문의 피인용 횟수로 오해하기 쉽다. 특히 5번째 논문의 9회를 보고 h-index를 9로 정하면 '최소 9편이 각각 9회 이상 인용'되어야 한다는 조건을 놓치게 된다.

다시 나오면 이렇게 판단

피인용 횟수를 내림차순으로 정렬한 뒤 각 순위 r에서 $c_r \geq r$인지 확인한다. 조건을 마지막으로 만족하는 순위가 h-index이며, 바로 다음 순위가 조건을 충족하지 않는지도 함께 확인한다.

암기 포인트

h-index는 '최소 h편이 각각 h회 이상 인용'되는 가장 큰 h이다.

NoteBase

문 20. 다음 중 적합성에 관한 황(Huang)과 소겔(Soergel)의 관점으로 가장 옳은 것은?

① 주제적 적합성은 이용자의 질의와 문헌에 나타난 핵심 용어의 단순 일치 여부로만 판단해야 하며, 이것이 객관적 적합성의 핵심이다.

② 이용자가 아동 자폐증의 치료법에 대한 정보를 원할 때, 어른 자폐증의 증상에 대한 문헌은 주제적으로 전혀 관련이 없으므로 부적합하다고 판단해야 한다.

③ 주제적 적합성은 '원인-결과', '비교-대조', '방법-해결' 등 다양한 관계 유형으로 세분화될 수 있으며, 정보시스템은 이러한 다차원적 관계를 고려하여 검색 결과를 제공할 필요가 있다.

④ 이용자의 주관적 판단보다는 시스템이 정해진 알고리즘에 따라 판단하는 객관적 주제 일치도가 적합성 평가의 최종 기준이 되어야 한다.

정답 3번

출제 포인트

황과 소겔이 주제적 적합성을 단순한 용어 일치가 아니라, 이용자의 정보요구와 문헌이 맺는 여러 기능적·의미적 관계를 포괄하는 개념으로 본다는 점을 묻는 문제이다.

핵심 개념

주제적 적합성은 문헌이 이용자의 주제나 과업에 어떤 방식으로 도움을 주는지를 나타내는 관계이다. 황과 소겔은 이를 단순 주제 일치에 한정하지 않고 직접적 관계, 간접적 관계, 맥락 제공, 비교 관계 등으로 세분하여 파악한다. 따라서 검색시스템도 원인과 결과, 문제와 해결, 비교와 대조처럼 개념 사이의 관계와 문헌의 기능적 역할을 고려할 필요가 있다.

선지별 해설

① X — '핵심 용어의 단순 일치 여부로만 판단해야 한다'는 부분이 틀렸다. 올바르게는 주제적 적합성을 단순한 용어 일치보다 넓게 보고, 문헌이 이용자의 주제와 맺는 직접·간접·맥락·비교 등의 관계까지 고려해야 한다.
② X — '주제적으로 전혀 관련이 없으므로 부적합하다'는 단정이 틀렸다. 아동 자폐증의 치료법과 어른 자폐증의 증상은 직접적인 정답 관계는 아닐 수 있지만, 자폐증의 발달 과정이나 비교·맥락을 제공하는 간접적 관련 정보가 될 수 있다.
③ O — 주제적 적합성을 원인과 결과, 비교와 대조, 방법과 해결 등 여러 관계 유형으로 세분하고, 검색시스템이 이러한 다차원적 관계를 반영해야 한다는 설명은 황과 소겔의 관점에 부합한다.
④ X — '객관적 주제 일치도가 적합성 평가의 최종 기준'이라는 부분이 틀렸다. 올바르게는 이용자의 정보요구와 과업, 문헌이 제공하는 기능적 역할을 함께 고려해야 하며, 시스템의 단순한 알고리즘적 주제 일치만을 최종 기준으로 삼지 않는다.

NoteBase

오답 포인트

①과 ④는 '객관적'이라는 말로 신뢰감을 주지만 실제로는 주제적 적합성을 단순 용어 일치로 축소한다. ②도 직접 답을 주지 않는 문헌을 곧바로 무관하다고 처리하여 간접·맥락·비교 관계를 놓치게 한다.

다시 나오면 이렇게 판단

'단순 일치만', '전혀 관련 없음', '최종 기준'처럼 적합성을 하나의 기준으로 제한하는 표현은 먼저 의심한다. 문헌이 원인, 결과, 맥락, 비교, 방법, 해결 등 어떤 역할로 이용자의 정보요구에 기여하는지를 인정하는 선지를 고른다.

암기 포인트

황·소겔의 주제적 적합성은 '단어 하나의 일치'가 아니라 '관계와 역할의 다차원성'이다.

NoteBase

문 21. 다음 중 자연어 색인 방식과 통제어 색인 방식의 비교 설명으로 가장 옳지 않은 것은?

① 통제어 색인은 시소러스 등을 활용하여 용어를 통제하므로 재현율은 낮아지지만, 정확률은 항상 자연어 색인보다 우수하게 나타나는 경향이 있다.

② 자연어 색인은 문헌에 사용된 용어를 그대로 색인어로 사용하므로 신조어 표현이 용이하나, 동음이의어나 동의어 문제가 발생하여 검색 효율이 낮아질 수 있다.

③ 통제어 색인은 색인 전문가의 용어 선택 과정이 필요하며, 구축된 통제어휘집의 최신성 유지를 위한 지속적인 관리가 요구된다.

④ 자연어 색인은 색인 과정이 비교적 단순하고 비용이 저렴할 수 있으나, 검색 시 이용자가 원하는 정보를 얻기 위해 다양한 동의어를 직접 고려해야 하는 부담이 있다.

정답 1번

출제 포인트

자연어 색인과 통제어 색인이 색인어를 선택하는 방식, 검색 성능, 구축·관리 비용에서 어떻게 다른지를 묻는 문제이다.

핵심 개념

자연어 색인: 문헌의 제목·초록·본문 등에 실제로 쓰인 용어를 그대로 색인어로 사용하는 방식이다. 신조어나 전문적인 표현을 빠르게 반영할 수 있지만, 동의어·이형어·동음이의어가 통제되지 않아 검색 누락이나 불필요한 검색 결과가 생길 수 있다.
통제어 색인: 시소러스나 주제명표처럼 미리 정한 표준 용어를 색인어로 사용하는 방식이다. 동의어를 하나의 대표어로 모으고 동음이의어를 구별하여 검색의 일관성을 높이지만, 전문적인 색인 작업과 통제어휘의 지속적인 갱신이 필요하다.

NoteBase

선지별 해설

① X — '재현율은 낮아지지만'과 '정확률은 항상 자연어 색인보다 우수하다'는 단정이 잘못되었다. 통제어 색인은 동의어를 대표어로 통합해 관련 문헌의 검색 누락을 줄일 수 있으므로 재현율 향상에 기여할 수 있고, 동음이의어를 구별해 정확률 향상에도 기여할 수 있다. 다만 실제 재현율과 정확률은 통제어휘의 품질, 최신성, 색인의 적절성, 검색 방식에 따라 달라지므로 항상 우수하다고 할 수 없다. 올바르게는 '통제어 색인은 용어 통제를 통해 재현율과 정확률을 높이는 데 기여할 수 있으나, 항상 자연어 색인보다 우수한 것은 아니다'이다.
② O — 자연어 색인은 문헌에 등장한 표현을 그대로 사용하므로 신조어를 즉시 반영하기 쉽다. 반면 동의어가 서로 다른 색인어로 흩어지면 관련 문헌이 누락될 수 있고, 동음이의어가 구별되지 않으면 불필요한 문헌까지 검색될 수 있다.
③ O — 통제어 색인은 문헌의 주제를 표준 용어에 대응시키는 판단이 필요하므로 색인 전문가의 개입이 요구될 수 있다. 또한 새 개념과 용어를 반영하고 용어 간 관계를 정비하려면 통제어휘집을 계속 관리해야 한다.
④ O — 자연어 색인은 별도의 통제어휘를 구축하고 적용하는 부담이 적어 색인 과정이 비교적 단순하고 비용이 낮을 수 있다. 대신 검색자는 동의어·이형어 등을 직접 조합하여 검색해야 하므로 검색 부담이 커질 수 있다.

오답 포인트

①은 통제어 색인이 용어를 제한한다는 점을 곧바로 '재현율 저하'로 연결하고, 정확률에는 '항상'이라는 절대 표현을 붙인 것이 함정이다. 통제어의 대표어 통합은 동의어 때문에 생기는 검색 누락을 줄일 수 있으며, 검색 성능은 어느 방식이든 조건에 따라 달라진다.

다시 나오면 이렇게 판단

자연어는 '원문 표현·신조어 반영·낮은 구축 부담'과 '동의어·동음이의어 문제'를 연결한다. 통제어는 '표준화·동의어 통합·의미 구별'과 '전문가 색인·어휘집 유지 비용'을 연결한다. 재현율이나 정확률에 '항상', '반드시'가 붙으면 조건에 따른 예외가 없는지 확인한다.

암기 포인트

자연어는 빠르고 자유롭지만 용어가 흩어지고, 통제어는 용어를 모아 주지만 구축과 갱신이 필요하다.

NoteBase

문 22. 정보격차에 영향을 주는 사회적 요인들에 대한 대표적인 이론으로 혁신확산(Diffusion of Innovations) 이론을 들 수 있다. 아래 보기 중 혁신확산이론에서 성공여부를 결정하는 혁신의 일반적인 속성만을 포함하고 있는 것은?

<보기>
ㄱ. 기능성(functionality)
ㄴ. 복잡성(complexity)
ㄷ. 상대적 이점(relative advantage)
ㄹ. 적합성(compatibility)
ㅁ. 통합성(integrity)

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

정답 3번

출제 포인트

혁신확산이론에서 혁신의 채택과 확산에 영향을 주는 일반적 속성을 구별하는 문제이다.

핵심 정리

혁신확산이론의 다섯 가지 일반적 속성은 상대적 이점, 적합성, 복잡성, 시험 가능성, 관찰 가능성이다. 상대적 이점·적합성·시험 가능성·관찰 가능성이 클수록 채택이 쉬워지고, 복잡성은 낮게 인식될수록 채택이 쉬워진다. 기능성과 통합성은 이 다섯 가지 속성에 포함되지 않는다.

선지별 해설

① X — ㄴ 복잡성과 ㄷ 상대적 이점은 해당하지만, ㄱ 기능성은 혁신의 다섯 가지 일반적 속성이 아니다. ㄱ 기능성을 ㄹ 적합성으로 바꾸면 ㄴ·ㄷ·ㄹ의 올바른 조합이 된다.
② X — ㄷ 상대적 이점과 ㄹ 적합성은 해당하지만, ㅁ 통합성은 다섯 가지 일반적 속성이 아니다. ㅁ 통합성을 ㄴ 복잡성으로 바꾸면 ㄴ·ㄷ·ㄹ의 올바른 조합이 된다.
③ O — ㄴ 복잡성, ㄷ 상대적 이점, ㄹ 적합성은 모두 혁신의 일반적 속성에 포함된다.
④ X — ㄷ 상대적 이점만 해당하고, ㄱ 기능성과 ㅁ 통합성은 해당하지 않는다. 올바른 조합은 ㄴ 복잡성·ㄷ 상대적 이점·ㄹ 적합성이다.

NoteBase

출제자가 노린 함정

기능성과 통합성은 혁신의 장점이나 완성도를 설명하는 말처럼 보여 선택하기 쉽다. 그러나 이론에서 정형화된 속성 명칭은 상대적 이점, 적합성, 복잡성, 시험 가능성, 관찰 가능성이다.

다시 나오면 이렇게 판단

보기에서 먼저 다섯 속성의 정식 명칭과 일치하는 항목만 남긴다. 이 문항에서는 ㄴ·ㄷ·ㄹ이 정식 속성이므로 이 셋을 모두 포함한 3번으로 판단한다.

암기 포인트

혁신의 다섯 속성은 상대적 이점·적합성·복잡성·시험 가능성·관찰 가능성이다.

NoteBase

문 23. 다음은 정보행위 관련 개념들의 관계를 나타낸 것이다. 이 관계에 대한 설명으로 가장 옳은 것은? (단, A ⊃ B는 A가 B를 포함한다는 의미이다)

정보행위(Information Behavior)
⊃ 정보추구행위(Information Seeking Behavior)
⊃ 정보탐색행위(Information Searching Behavior)

① 정보탐색행위는 정보추구행위보다 더 넓은 개념으로, 정보시스템과의 상호작용 없이 이루어지는 모든 정보 획득 활동을 포함한다.

② 정보행위는 개인이 정보를 의식적으로 찾아 나서는 능동적인 행동만을 지칭하며, 우연히 정보를 찾는 수동적인 경우는 제외된다.

③ 정보추구행위는 특정한 목적 달성을 위해 의식적으로 정보를 찾는 행위를 의미하며, 정보탐색행위는 그중에서도 정보시스템을 활용한 정보 추구 활동을 구체적으로 지칭한다.

④ 위 세 가지 개념은 완전히 독립적이어서, 하나의 정보 관련 활동이 동시에 여러 범주에 속하는 것은 불가능하다.

정답 3번

출제 포인트

정보행위, 정보추구행위, 정보탐색행위의 포함관계와 각 개념의 범위를 구별하는 문제이다.

핵심 개념

정보행위는 사람이 정보와 관련하여 하는 가장 넓은 범위의 활동으로, 의도적인 탐색뿐 아니라 우연한 정보 획득과 수동적인 정보 수용도 포함한다.
정보추구행위는 필요나 목적을 충족하기 위해 의식적으로 정보를 찾는 활동이다.
정보탐색행위는 정보추구행위 가운데 정보검색시스템 등 정보시스템과 상호작용하며 수행하는 구체적인 탐색 활동이다.
따라서 포함관계는 정보행위 ⊃ 정보추구행위 ⊃ 정보탐색행위이다.

NoteBase

선지별 해설

① X
정보탐색행위가 정보추구행위보다 '더 넓은 개념'이라는 부분과 '정보시스템과의 상호작용 없이 이루어지는 모든 정보 획득 활동을 포함한다'는 부분이 틀리다. 올바르게는 정보탐색행위가 정보추구행위에 포함되는 더 좁은 개념이며, 정보시스템과의 상호작용을 통한 구체적 탐색 활동을 가리킨다.

② X
정보행위가 '능동적인 행동만'을 지칭하고 수동적인 경우를 제외한다는 부분이 틀리다. 올바르게는 정보행위가 의도적인 정보 추구뿐 아니라 우연히 정보를 접하거나 수동적으로 정보를 받아들이는 경우까지 포괄한다.

③ O
정보추구행위는 목적을 가지고 의식적으로 정보를 찾는 행위이고, 정보탐색행위는 그중 정보시스템을 활용하여 구체적으로 탐색하는 활동이므로 제시된 포함관계와 일치한다.

④ X
세 개념이 '완전히 독립적'이라는 부분이 틀리다. 올바르게는 세 개념이 포함관계에 있으므로 하나의 정보탐색 활동은 동시에 정보탐색행위, 정보추구행위, 정보행위에 모두 속할 수 있다.

오답 포인트

①은 정보탐색행위와 정보추구행위의 범위를 거꾸로 제시하고, ②는 가장 넓은 정보행위를 의도적·능동적 행위로만 축소한다. '탐색'이라는 말이 일상적으로 넓게 느껴지더라도 이론상 정보탐색행위가 가장 좁은 개념이라는 점이 핵심 함정이다.

다시 나오면 이렇게 판단

먼저 의도 여부와 정보시스템 상호작용 여부를 확인한다. 우연하거나 수동적인 정보 획득까지 포함하면 정보행위, 목적을 가지고 의식적으로 찾으면 정보추구행위, 그 과정에서 정보시스템과 구체적으로 상호작용하면 정보탐색행위로 판단한다.

암기 포인트

행위가 가장 넓고, 목적 있는 추구가 그 안에 있으며, 시스템을 통한 탐색이 가장 좁다: 정보행위 ⊃ 정보추구행위 ⊃ 정보탐색행위.

NoteBase

문 24. OSI(Open Systems Interconnection) 모델은 ISO(국제 표준기구)에서 표준화된 네트워크 구조를 정의하기 위해서 제공한 기본 참조 모델로 7개의 계층으로 구성되어 있다. 다음 중 제3계층부터 제5계층을 순서대로 나열한 것은?

① 데이터 링크 계층 - 네트워크 계층 - 전송 계층

② 네트워크 계층 - 전송 계층 - 세션 계층

③ 전송 계층 - 세션 계층 - 표현 계층

④ 세션 계층 - 표현 계층 - 응용 계층

정답 2번

출제 포인트

OSI 7계층에서 제3계층부터 제5계층까지의 명칭과 배열 순서를 묻는 문제이다.

올바른 순서

OSI 계층은 아래에서부터 물리 계층(1) → 데이터 링크 계층(2) → 네트워크 계층(3) → 전송 계층(4) → 세션 계층(5) → 표현 계층(6) → 응용 계층(7) 순서이다.
따라서 제3계층부터 제5계층은 네트워크 계층 → 전송 계층 → 세션 계층이다.
왜 이 순서인가: 계층 번호가 하나씩 올라갈 때 네트워크 계층의 경로 선택·패킷 전달 기능 위에 전송 계층의 종단 간 전송 기능이 놓이고, 그 위에 세션 계층의 통신 세션 관리 기능이 놓이기 때문이다.

선지별 해설

① X — 데이터 링크 계층은 제2계층이므로 제3계층부터 시작하는 배열이 아니다. 틀린 부분은 '데이터 링크 계층 - 네트워크 계층 - 전송 계층'이며, 올바르게는 '네트워크 계층 - 전송 계층 - 세션 계층'이다.
② O — 네트워크 계층은 제3계층, 전송 계층은 제4계층, 세션 계층은 제5계층이므로 제3계층부터 제5계층까지를 정확히 나열한다.
③ X — 전송 계층, 세션 계층, 표현 계층은 각각 제4·5·6계층이다. 틀린 부분은 제4계층인 전송 계층부터 시작한 것이며, 올바르게는 '네트워크 계층 - 전송 계층 - 세션 계층'이다.
④ X — 세션 계층, 표현 계층, 응용 계층은 각각 제5·6·7계층이다. 틀린 부분은 제5계층인 세션 계층부터 시작한 것이며, 올바르게는 '네트워크 계층 - 전송 계층 - 세션 계층'이다.

오답 포인트

각 선지는 연속된 세 계층을 제시하지만 시작 계층만 다르다. '제3계층부터'라는 조건을 놓치면 제2계층이나 제4·5계층에서 시작하는 배열을 고르기 쉽다.

NoteBase

다시 나오면 이렇게 판단

먼저 OSI 7계층을 아래에서부터 번호와 함께 대응한 뒤, 문제에서 요구한 시작 번호와 끝 번호만 잘라 판단한다. 제3계층부터 제5계층이라면 3·4·5에 해당하는 네트워크·전송·세션을 고른다.

암기 포인트

OSI 아래→위: 물·데·네·전·세·표·응, 3·4·5는 네·전·세이다.

NoteBase

문 25. 다음 중 정보 검색 결과 평가 시, 재현율(Recall)을 높이기 위한 전략으로 가장 적절한 것은?

① 검색어의 동의어, 유의어를 추가하여 검색한다.

② 검색 범위를 특정 필드(예: 제목)로 제한한다.

③ 불리언 연산자 AND를 사용하여 검색어를 조합한다.

④ 보다 구체적이고 상세한 검색어를 사용한다.

정답 1번

출제 포인트

이 문제는 정보 검색의 재현율을 높이는 검색 확장 전략과 검색 범위를 좁히는 전략을 구별하는 문제이다.

핵심 개념

재현율은 전체 적합 문헌 가운데 실제로 검색된 적합 문헌의 비율이다. 재현율을 높이려면 동의어·유의어를 추가하거나 OR로 연결하여 검색 범위를 넓힌다. 반면 특정 필드 제한, AND 결합, 구체적인 검색어 사용은 검색 범위를 좁혀 일반적으로 정확률을 높이는 방향으로 작용한다.

선지별 해설

① O — 검색어의 동의어와 유의어를 추가하면 같은 개념을 다른 표현으로 기록한 문헌까지 검색할 수 있으므로 누락이 줄고 재현율이 높아진다.
② X — 틀린 부분은 '특정 필드로 제한한다'이다. 제목처럼 특정 필드만 검색하면 다른 필드에 관련 용어가 있는 문헌이 제외된다. 재현율을 높이려면 검색 필드를 제한하기보다 필요한 범위에서 넓혀야 한다.
③ X — 틀린 부분은 'AND를 사용한다'이다. AND는 모든 검색어를 함께 포함한 문헌만 남겨 검색 결과를 좁힌다. 재현율을 높이는 데에는 동의어·유의어를 OR로 연결하는 방식이 적절하다.
④ X — 틀린 부분은 '보다 구체적이고 상세한 검색어'이다. 검색어가 구체적일수록 조건을 충족하는 문헌 수가 줄어 관련 문헌이 누락될 수 있다. 재현율을 높이려면 포괄적인 용어와 관련 표현을 함께 사용하여 검색 범위를 넓혀야 한다.

오답 포인트

②·③·④는 검색 결과의 관련성을 높이는 데 도움이 될 수 있어 좋아 보이지만, 공통적으로 검색 범위를 좁히는 전략이다. 이 문항은 정확률이 아니라 '관련 문헌을 얼마나 빠짐없이 찾는가'를 나타내는 재현율을 묻는다.

다시 나오면 이렇게 판단

검색 범위를 넓혀 누락을 줄이는 전략이면 재현율 향상, 검색 조건을 강화하여 불필요한 결과를 줄이는 전략이면 정확률 향상으로 판단한다. 동의어·유의어·OR는 전자에, 필드 제한·AND·구체화는 후자에 해당한다.

NoteBase

암기 포인트

재현율은 넓고 빠짐없이, 정확률은 좁고 정확하게이다.

NoteBase