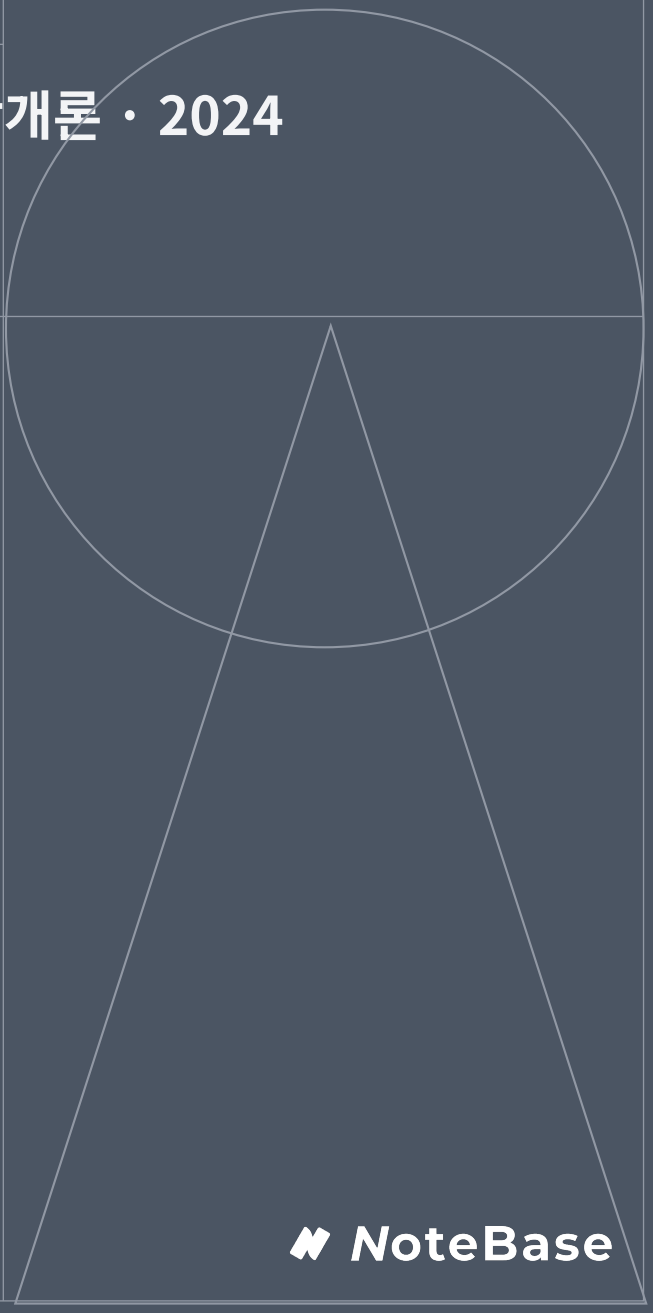
기출OX 해설집

공무원/법원직/9급 · 정보학개론 · 2024



NoteBase

기출OX 해설집

공무원/법원직/9급 · 정보학개론 · 2024

공무원/법원직/9급 · 정보학개론 · 2024

문 1. 다음 <보기>에서 다크웹(Dark Web)에 대한 설명으로 옳은 것을 모두 고른 것은?

<보기>
ㄱ. 본래 미 해군 연구소에서 익명성을 목적으로 만들어진 기술이다.
ㄴ. 익명성 보장은 물론 IP주소 추적이 불가능하도록 고안된 인터넷 영역이다.
ㄷ. 일반적인 검색엔진으로는 찾을 수 없기 때문에 해킹으로 얻은 개인정보, 경쟁사의 영업비밀 등 주로 불법적인 정보가 거래되는 웹이다.
ㄹ. 웹페이지를 찾아다니는 웹 크롤러에 의해 수집되지 않아 검색 등으로는 접근이 어려운 웹이다.

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

정답 4번

출제 포인트

다크웹의 기원과 익명성 기술, 검색엔진 비노출성, 불법 정보 유통 위험을 종합하여 판단하는 문제이다.

핵심 개념

다크웹은 일반 검색엔진에 색인되지 않으며 Tor와 같은 전용 네트워크·도구를 통해 접근하는 웹 영역이다.
ㄱ: 양파 라우팅과 Tor 기술은 미국 해군연구소의 익명 통신 연구에서 출발했으므로 옳다.
ㄴ: 여러 중계 지점을 거치고 통신을 암호화하여 이용자와 서비스의 실제 IP 주소를 숨기도록 설계되었으므로 옳다.
ㄷ: 일반 검색엔진으로 찾기 어렵고 익명성이 높아 탈취 개인정보나 영업비밀 등 불법 정보가 거래되는 공간으로 악용되므로 옳다. 다만 익명 제보나 검열 회피 같은 합법적 용도도 존재한다.
ㄹ: 웹 크롤러가 수집·색인하지 않는 영역이어서 일반적인 검색만으로 접근하기 어려우므로 옳다.

NoteBase

선지별 해설

① X — ㄱ과 ㄴ은 옳지만, 옳은 설명인 ㄷ과 ㄹ을 빠뜨렸다. 올바르게는 'ㄱ, ㄴ, ㄷ, ㄹ'이다.
② X — ㄱ, ㄴ, ㄷ은 옳지만, 검색엔진의 웹 크롤러에 수집되지 않아 접근이 어렵다는 ㄹ도 옳다. 올바르게는 'ㄱ, ㄴ, ㄷ, ㄹ'이다.
③ X — ㄱ, ㄴ, ㄹ은 옳지만, 익명성과 검색 비노출성이 불법 정보 거래에 악용된다는 ㄷ도 옳다. 올바르게는 'ㄱ, ㄴ, ㄷ, ㄹ'이다.
④ O — ㄱ은 기술의 기원, ㄴ은 익명성 설계, ㄷ은 불법 거래 악용, ㄹ은 검색엔진 비색인 특성을 각각 바르게 설명하므로 모두 옳다.

오답 포인트

ㄹ을 딥웹에만 해당하는 설명으로 보아 제외하기 쉽다. 다크웹은 딥웹의 일부이므로 일반 검색엔진에 색인되지 않는다는 특징을 함께 가진다. 또한 다크웹에는 합법적 용도도 있지만, ㄷ은 불법 정보 유통에 악용되는 대표적 위험을 설명한 문장이다.

다시 나오면 이렇게 판단

다크웹 문항에서는 '일반 검색엔진 비색인'이라는 딥웹의 공통 속성에 '전용 네트워크를 통한 접근'과 '익명성'이 더해지는지를 확인한다. 기원은 미국 해군연구소의 양파 라우팅 연구, 기술적 핵심은 실제 IP 주소 은닉과 다중 중계이다.

암기 포인트

다크웹은 '비색인 + 전용 접근 + 익명성'으로 정리한다.

NoteBase

문 2. 정보검색시스템의 성능 평가를 위한 대표적인 척도로 재현율과 정확률이 있다. 재현율과 정확률은 일반적으로 반비례 관계에 있으며 시스템의 검색 효율을 판단하기 위해서는 함께 사용된다. 하지만 두 개 이상의 시스템 성능을 비교할 경우 어느 시스템 성능이 더 나은지를 판단하기 어렵기 때문에 F-척도와 같은 단일 척도가 개발되기도 했다.
가상적으로 어떤 검색시스템의 재현율이 50%, 정확률이 75%라고 한다면 해당 시스템의 F-척도값은 다음 중 어떤 것인가?

① 0.45

② 0.50

③ 0.60

④ 0.75

정답 3번

출제 포인트

재현율과 정확률의 조화평균인 F-척도를 계산하는 문제이다.

조건 정리

주어진 값: 재현율 $R = 0.50$, 정확률 $P = 0.75$이다.
구해야 할 값: 재현율과 정확률을 함께 반영한 F-척도이다.
주의할 조건: 백분율을 각각 0.50과 0.75로 바꾸어 계산한다.

사용할 공식

$F = \frac{2PR}{P+R}$
F-척도는 재현율과 정확률 가운데 한 값만 보거나 단순평균하지 않고, 두 값의 조화평균으로 검색 성능을 종합하기 때문에 이 공식을 사용한다.

계산 과정

$F = \frac{2 \times 0.75 \times 0.50}{0.75 + 0.50} = \frac{0.75}{1.25} = 0.60$
따라서 F-척도값은 0.60이다.

선지 판단

① X — 0.45는 공식에 대입한 결과와 일치하지 않는다. 올바른 값은 0.60이다.
② X — 0.50은 주어진 재현율 자체이며 F-척도값이 아니다. 올바른 값은 0.60이다.
③ O — 조화평균 공식으로 계산한 F-척도값이 0.60이므로 옳다.
④ X — 0.75는 주어진 정확률 자체이며 F-척도값이 아니다. 올바른 값은 0.60이다.

NoteBase

오답 포인트

재현율 0.50이나 정확률 0.75 중 하나를 그대로 고르는 것이 함정이다. 두 비율을 함께 반영하려면 산술평균이 아니라 조화평균 공식을 적용해야 한다.

다시 나오면 이렇게 판단

재현율과 정확률이 주어지고 F-척도를 묻는 경우에는 백분율을 소수로 바꾼 뒤 $2PR/(P+R)$에 대입하고, 결과가 두 값 사이에 있는지 검산한다.

암기 포인트

F-척도는 재현율과 정확률의 조화평균이며, 공식은 $F = 2PR/(P+R)$이다.

NoteBase

문 3. 디지털도서관의 발전에 따라 사서의 도서관 자료와 관련된 저작권에 대한 이해와 적용이 요구된다. 공정이용 (Fair Use)은 저작권자의 배타적인 권리를 제한하는 규정으로 1976년 미국 저작권법에 의해 성문화된 바 있다. 공정이용은 저작권법의 침해를 받지 않는 언론 보도, 교육, 연구 등을 목적으로 하는 저작물의 이용을 뜻하며 전통적으로 도서관 자료 서비스의 근거가 되어 왔다.

다음 중 공정이용을 결정하는 요건으로 보기 힘든 것은?

① 사용의 목적과 성격

② 저작권이 있는 저작물의 성격

③ 공정이용과 관련된 정보의 분량

④ 저작자의 허락 또는 승인

정답 4번

쟁점

이 문제는 미국 저작권법상 공정이용을 판단하는 네 가지 요소에 관한 문제이다. 핵심은 법정 판단 요소와 저작권자의 사전 허락을 구별하는 것이다.

용어 풀이

공정이용은 비평·논평·보도·교육·학술·연구 등의 목적을 포함하여, 저작권자의 허락이 없더라도 구체적 사정에 따라 저작권 침해가 되지 않을 수 있는 이용이다.

핵심 구분

미국 저작권법 제107조는 ① 이용의 목적과 성격, ② 저작물의 성격, ③ 이용된 부분의 양과 질적 중요성, ④ 그 이용이 저작물의 잠재적 시장이나 가치에 미치는 영향을 함께 고려하도록 한다. 저작권자의 허락 또는 승인은 이 네 가지 판단 요소가 아니라, 권리자로부터 이용 권한을 얻는 별개의 방식이다.

선지별 해설

① O — 사용의 목적과 성격은 공정이용의 첫 번째 판단 요소이다. 상업적 이용인지 비영리 교육 목적의 이용인지 등이 고려된다.

② O — 저작권이 있는 저작물의 성격은 두 번째 판단 요소이다. 저작물이 사실 중심인지 창작성이 강한지 등의 성격을 살핀다.

③ O — 공정이용과 관련된 정보의 분량은 이용된 부분의 양과 질적 중요성을 가리키는 것으로 볼 수 있으므로 판단 요소에 해당한다. 더 정확한 표현은 '저작물 전체에서 이용된 부분의 양과 질적 중요성'이다.

④ X — '저작자의 허락 또는 승인'은 공정이용의 네 가지 판단 요소에 포함되지 않는다. 틀린 표현은 '저작자의 허락 또는 승인'이며, 올바른 네 번째 요소는 '그 이용이 저작물의 잠재적 시장이나 가치에 미치는 영향'이다.

NoteBase

오답 포인트

③의 '정보의 분량'이라는 다소 부정확한 표현과 ④의 '허락 또는 승인'을 혼동하기 쉽다. ③은 이용 부분의 양과 질적 중요성을 뜻하는 법정 요소이지만, ④의 허락은 공정이용 판단과 별개의 이용 권한이다.

다시 나오면 이렇게 판단

공정이용 문항에서는 목적·저작물 성격·이용량과 중요성·시장 영향을 차례로 대조한다. 이 목록에 없는 허락·승인은 공정이용 요소가 아니라 별도의 권리 처리 방식으로 판단한다.

암기 포인트

공정이용의 네 요소는 '목적—저작물 성격—이용량과 중요성—시장 영향'이다. 허락은 네 요소 밖이다.

문 4. 벡터 스페이스 모델(Vector Space Model, VSM)에 관한 다음 설명 중 가장 옳지 않은 것은?

① 벡터 스페이스 모델은 문서와 쿼리를 벡터로 표현하여 유사도를 계산하는 모델이다.

② 벡터 스페이스 모델에서 문서와 쿼리는 단어의 빈도를 기반으로 벡터 공간에 위치하게 된다.

③ 벡터 스페이스 모델에서 문서와 쿼리 간의 유사도는 주로 코사인 유사도(Cosine Similarity)로 측정된다.

④ 벡터 스페이스 모델은 문서 간의 순서나 구조를 고려하여 문서의 의미를 파악하는 데 중점을 둔다.

정답 4번

출제 포인트

벡터 스페이스 모델이 문서와 질의를 어떤 방식으로 표현하고 비교하는지, 그리고 전통적인 모델이 단어의 순서와 문서 구조를 반영하는지를 묻는 문제이다.

핵심 개념

벡터 스페이스 모델은 문서와 질의를 용어별 가중치로 이루어진 벡터로 표현하고, 두 벡터의 방향이 얼마나 비슷한지를 이용하여 관련성을 계산하는 정보 검색 모델이다. 전통적인 벡터 스페이스 모델은 단어의 출현과 가중치에 초점을 두므로 단어의 순서나 문장·문서 구조는 직접 반영하지 않는다.

선지별 해설

① O — 문서와 쿼리를 같은 용어 공간의 벡터로 나타내면 두 벡터 사이의 유사도를 계산하여 관련 문서를 찾을 수 있다.
② O — 각 용어가 벡터의 차원이 되고, 문서와 쿼리의 단어 출현 빈도나 이를 보정한 가중치가 각 차원의 값이 된다. 따라서 단어 빈도를 바탕으로 벡터 공간에 표현된다는 설명은 옳다.
③ O — 코사인 유사도는 두 벡터가 이루는 각도를 바탕으로 방향의 유사성을 측정하므로 문서와 쿼리의 유사도 계산에 대표적으로 사용된다.
④ X — 틀린 부분은 '문서 간의 순서나 구조를 고려하여 문서의 의미를 파악하는 데 중점을 둔다'이다. 전통적인 벡터 스페이스 모델은 단어의 순서와 문서 구조를 무시하고, 용어의 출현 빈도와 가중치를 바탕으로 문서와 쿼리의 유사도를 계산한다고 고쳐야 한다.

오답 포인트

④는 문서를 수치 벡터로 표현한다는 사실을 문장의 순서와 구조까지 이해한다는 뜻으로 바꾼 선지이다. 벡터화 자체와 어순·구조의 반영은 별개이며, 전통적인 벡터 스페이스 모델은 이른바 단어 주머니 방식의 한계를 가진다.

다시 나오면 이렇게 판단

'용어별 벡터·빈도 또는 가중치·코사인 유사도'는 벡터 스페이스 모델의 핵심 특징으로 판단한다. 반대로 '어순·문장 구조·문맥을 직접 고려한다'는 표현은 전통적인 벡터 스페이스 모델의 특징이 아닌 것으로 판단한다.

NoteBase

암기 포인트

VSM은 용어를 벡터로, 관련성은 코사인으로 판단하며, 전통적 VSM은 어순과 구조를 직접 보지 않는다.

NoteBase

문 5. 다음 중 정보 사이클에서 분류, 편목, 색인 초록 작업을 수행하는 단계는 무엇인가?

① 정보생산

② 정보조직

③ 정보축적

④ 정보검색

정답 2번

출제 포인트

정보 사이클의 각 단계가 담당하는 작업을 구분하는 문제이다. 분류·편목·색인·초록은 정보를 일정한 체계로 정리하여 찾을 수 있게 만드는 정보조직 작업이다.

올바른 순서

대표적인 흐름은 정보생산 → 정보조직 → 정보축적 → 정보검색으로 파악할 수 있다.
정보생산: 새로운 정보가 만들어지는 단계이다.
정보조직: 분류·편목·색인·초록 등을 통해 정보를 체계적으로 정리하는 단계이다.
정보축적: 조직된 정보를 데이터베이스나 정보시스템 등에 저장하는 단계이다.
정보검색: 축적된 정보에서 이용자의 요구에 맞는 정보를 찾아내는 단계이다.
정보를 먼저 조직해야 체계적으로 축적하고 효율적으로 검색할 수 있다.

선지별 해설

① X — 정보생산은 새로운 지식이나 기록이 만들어지는 단계이다. 틀린 표현인 '정보생산'은 '정보조직'으로 고쳐야 한다.
② O — 정보조직은 분류·편목·색인·초록을 통해 정보의 내용과 형태를 분석하고 검색 가능한 질서를 부여하는 단계이다.
③ X — 정보축적은 조직된 정보를 저장하고 보존하는 단계이다. 틀린 표현인 '정보축적'은 '정보조직'으로 고쳐야 한다.
④ X — 정보검색은 축적된 정보에서 이용자가 원하는 정보를 찾아내는 단계이다. 틀린 표현인 '정보검색'은 '정보조직'으로 고쳐야 한다.

오답 포인트

분류·편목·색인·초록을 정보의 저장이나 검색 자체로 혼동하기 쉽다. 이 작업들은 저장 전에 정보를 일정한 기준으로 정리하고 접근점을 만드는 정보조직에 해당한다.

NoteBase

다시 나오면 이렇게 판단

업무의 동사를 기준으로 판단한다. ‘만든다’는 정보생산, ‘분류하고 기술한다’는 정보조직, ‘저장한다’는 정보축적, ‘찾아낸다’는 정보검색에 대응한다.

암기 포인트

분류·편목·색인·초록은 모두 정보를 정리하는 작업이므로 정보조직이다.

문 6. 용어-문헌 인접 행렬(Term-Document Adjacency Matrix)에 관한 다음 설명 중 가장 옳지 않은 것은?

① 용어-문헌 인접 행렬은 문서 컬렉션에서 각 문서에 등장하는 각 용어의 빈도를 행렬 형태로 나타낸 것이다.

② 행렬의 행은 문서를, 열은 용어를 나타내며, 셀의 값은 해당 문서에서 해당 용어의 빈도이다.

③ 용어-문헌 인접 행렬은 문서 간의 유사도를 계산하기 위한 기초 자료로 사용될 수 있다.

④ 용어-문헌 인접 행렬은 문서가 추가되거나 삭제될 때마다 행렬을 갱신해야 하는 단점이 있다.

정답 2번

출제 포인트

용어-문헌 인접 행렬에서 행과 열이 각각 무엇을 나타내는지, 그리고 이 행렬이 어떻게 활용·갱신되는지를 묻는 문제이다.

핵심 개념

용어-문헌 인접 행렬은 문서 집합에 포함된 용어와 문헌의 관계를 수치로 나타낸 행렬이다. 일반적인 구성은 행이 용어, 열이 문헌이며, 각 셀에는 해당 용어가 해당 문헌에 나타난 빈도나 이를 가공한 가중치가 들어간다. 각 문헌은 하나의 열벡터로 표현되므로 문헌 간 유사도 계산에도 활용할 수 있다.

선지별 해설

① O — 각 셀에 문헌별 용어 출현 빈도를 기록하므로 문서 컬렉션의 용어 분포를 행렬로 표현한 것이라는 설명은 옳다.
② X — 틀린 부분은 '행은 문서, 열은 용어'이다. 일반적인 용어-문헌 인접 행렬에서는 행이 용어이고 열이 문헌이며, 셀 값은 해당 용어가 해당 문헌에 출현한 빈도이다.
③ O — 각 문헌이 용어별 값으로 이루어진 벡터가 되므로 코사인 유사도 같은 방법으로 문헌 간 유사도를 계산하는 기초 자료가 될 수 있다.
④ O — 문헌이 추가되거나 삭제되면 문헌에 대응하는 열과 관련 값이 달라지므로 행렬을 갱신해야 한다는 설명은 옳다.

오답 포인트

②는 행과 열의 역할만 서로 바꾸어 제시한 선지이다. 셀 값에 대한 설명은 맞지만 축의 방향이 뒤바뀌었으므로 전체 선지는 옳지 않다.

다시 나오면 이렇게 판단

행렬 이름의 순서를 기준으로 첫 번째 대상인 용어를 행, 두 번째 대상인 문헌을 열에 대응시키고, 셀은 '그 용어가 그 문헌에 나타난 정도'로 판단한다.

NoteBase

암기 포인트

용어-문헌 행렬은 행=용어, 열=문헌, 셀=출현 빈도이다.

NoteBase

문 7. 윌슨(Wilson)의 정보행위모형에 관한 다음 설명 중 가장 옳지 않은 것은?

① 심리학 등 다른 분야의 이론을 접목하여 정보추구의 복잡한 맥락을 강조하였다.

② 스트레스.대처이론을 이용하여 왜 어떤 정보원은 다른 정보원보다 더 많이 이용되는지 설명하고자 하였다.

③ 사회학습이론을 이용하여 왜 어떤 목표는 성취하고 다른 목표는 성취하지 못하는지 설명하고자 하였다.

④ 이용자로 하여금 정보추구과정을 시작하게 하는 동기요인인 촉진 메카니즘을 제시하였다.

정답 2번

출제 포인트

이 문제는 윌슨의 정보행위모형에 도입된 심리학 이론과 각 이론이 설명하는 대상을 정확히 연결하는 문제이다.

핵심 정리

윌슨의 정보행위모형은 정보요구가 곧바로 정보추구로 이어진다고 보지 않고, 개인·사회적 맥락과 촉진 메커니즘을 함께 설명한다. 스트레스·대처이론은 정보요구가 정보추구행동을 촉발하는 조건을 설명하고, 위험·보상이론은 정보원을 선택하고 이용하는 양상을 설명하며, 사회학습이론의 자기효능감은 목표 달성 가능성과 관련된다.

선지별 해설

① O — 윌슨은 심리학 등 여러 분야의 이론을 도입하여 정보행위를 개인의 단순한 탐색 절차가 아니라 복합적인 맥락 속 행동으로 설명하였다.
② X — 틀린 표현은 '스트레스·대처이론을 이용하여 왜 어떤 정보원은 다른 정보원보다 더 많이 이용되는지 설명'한다는 부분이다. 정보원 이용의 차이는 위험·보상이론으로 설명하며, 스트레스·대처이론은 정보요구가 정보추구행동을 촉발하는 조건과 관련된다.
③ O — 사회학습이론의 핵심 요소인 자기효능감은 개인이 어떤 행동이나 목표를 성공적으로 수행할 수 있다고 믿는 정도이므로, 목표의 성취 여부를 설명하는 데 활용된다.
④ O — 윌슨은 정보요구가 실제 정보추구행동으로 이어지게 하는 동기적 장치로 촉진 메커니즘을 제시하였다.

출제자가 노린 함정

②는 윌슨 모형에 실제로 사용된 '스트레스·대처이론'이라는 명칭을 제시한 뒤 그 설명 대상을 '정보원 이용의 차이'로 바꾸었다. 이 설명 대상은 위험·보상이론에 해당한다.

다시 나오면 이렇게 판단

윌슨 모형의 이론을 '스트레스·대처 → 정보추구의 촉발', '위험·보상 → 정보원 선택·이용', '사회학습·자기효능감 → 목표 성취'로 연결하여 판단한다.

NoteBase

암기 포인트

촉발은 스트레스·대처, 정보원 선택은 위험·보상, 목표 성취는 자기효능감이다.

NoteBase

문 8. 인터넷에 연결된 모든 기기는 고유의 IP 주소를 가지고 있다. IP 주소 체계에 관한 다음 설명 중 가장 옳지 않은 것은?

① 현재 사용되고 있는 IPv4는 점(.)으로 구분된 4개의 옥텟(Octet)으로 구성된 32비트 주소 체계로 표현된다.

② 차세대 인터넷 주소 체계인 IPv6는 IPv4의 제한된 주소 공간을 해소하기 위해 개발되었고, 256비트의 길이로 주소 공간이 확대되었다.

③ 현재 IPv4와 IPv6가 혼용되고 있다.

④ IPv6는 인터넷 표준으로 채택되었다.

정답 2번

출제 포인트

IPv4와 IPv6의 주소 길이, 표기 방식, 도입 목적을 구분하는 문제이다.

핵심 개념

IPv4는 32비트 주소를 8비트씩 4개의 옥텟으로 나누어 점으로 구분해 표기하는 주소 체계이다.
IPv6는 IPv4의 주소 부족 문제를 해결하기 위해 마련된 128비트 주소 체계이다.

선지별 해설

① O — IPv4 주소는 총 32비트이며, 8비트 단위인 4개의 옥텟을 점으로 구분해 표현한다.
② X — 틀린 표현은 '256비트의 길이'이다. IPv6는 IPv4의 제한된 주소 공간을 확대하기 위해 개발되었지만, 주소 길이는 256비트가 아니라 128비트이다.
③ O — IPv6로의 전환이 한 번에 완료되는 것은 아니므로 현재 IPv4와 IPv6가 함께 사용된다.
④ O — IPv6는 표준화된 차세대 인터넷 프로토콜 주소 체계이다.

오답 포인트

②의 '256비트'는 주소 공간이 크게 확대된다는 점을 과장해 바꾼 표현이다. IPv4의 32비트와 IPv6의 128비트를 정확히 구분해야 한다.

다시 나오면 이렇게 판단

주소 체계 문제에서는 먼저 비트 수를 확인한다. IPv4는 32비트·4옥텟·점 표기이고, IPv6는 128비트라는 기준으로 선지를 판별한다.

암기 포인트

IPv4는 32비트, IPv6는 128비트이다.

NoteBase

문 9. 디지털도서관에서는 도서관 내외의 정보자료를 이용자들이 보다 손쉽게 이용할 수 있도록 다양한 이용자 인증 방법을 활용하고 있다. 다음 중 오픈소스 기반의 이용자 인증 솔루션으로 기관 간의 인증을 통해 기관이 자신의 이용자가 정보를 안정적으로 교환할 수 있도록 하는 프로토콜은 무엇인가?

① LDAP(Lightweight Directory Access Protocol)

② DRM(Digital Rights Management)

③ PKI(Public Key Infrastructure)

④ Shibboleth

정답 4번

출제 포인트

기관 간 연계 인증을 지원하는 Shibboleth와 디렉터리 접근, 저작권 관리, 공개키 기반 보안을 담당하는 기술을 구별하는 문제이다.

핵심 개념

Shibboleth는 오픈소스 기반의 연합 인증 솔루션으로, 이용자의 소속 기관이 인증을 담당하고 다른 기관의 서비스가 그 인증 정보를 신뢰하여 접근을 허용하게 한다. SAML 기반으로 기관 간 인증 정보를 교환하며 통합 인증을 지원한다.
LDAP는 디렉터리 정보 조회·수정을 위한 접근 프로토콜이고, DRM은 디지털 콘텐츠의 이용 권한을 통제하는 기술이며, PKI는 공개키·인증서를 이용해 신원 확인과 암호화 기반을 제공하는 체계이다.

선지별 해설

① X — LDAP는 사용자 계정 등 디렉터리 정보를 조회하고 관리하기 위한 접근 프로토콜이다. 틀린 부분은 이를 기관 간 연합 인증 솔루션으로 본 점이며, 제시된 설명에 해당하는 것은 Shibboleth이다.
② X — DRM은 디지털 콘텐츠의 복제·열람·배포 등 이용 권한을 관리하는 기술이다. 틀린 부분은 이를 기관 간 이용자 인증 기술로 본 점이며, 올바른 개념은 디지털 저작권 관리이다.
③ X — PKI는 공개키, 개인키, 전자 인증서를 바탕으로 인증·암호화·전자서명을 지원하는 기반 체계이다. 틀린 부분은 이를 오픈소스 연합 인증 솔루션으로 본 점이며, 제시된 설명에 해당하는 것은 Shibboleth이다.
④ O — Shibboleth는 기관별 인증을 상호 신뢰하도록 연결하는 오픈소스 연합 인증 솔루션이다. 이용자는 소속 기관에서 인증받고 다른 참여 기관의 정보자원에 접근할 수 있으므로 문제의 설명과 일치한다.

오답 포인트

LDAP와 PKI도 인증 과정에 활용될 수 있어 혼동하기 쉽다. 그러나 문제의 결정적 표현은 '오픈소스 기반'과 '기관 간 인증'이다. 이 두 표현은 개별 디렉터리 조회나 인증서 기반 보안이 아니라 연합 인증을 구현하는 Shibboleth를 가리킨다.

NoteBase

다시 나오면 이렇게 판단

기관 간 인증·연합 인증·소속 기관의 인증 정보 공유가 나오면 Shibboleth를 판단한다. 디렉터리 조회는 LDAP, 콘텐츠 이용 통제는 DRM, 공개키와 인증서는 PKI로 구분한다.

암기 포인트

Shibboleth는 기관 간 연합 인증, LDAP는 디렉터리, DRM은 콘텐츠 권한, PKI는 공개키·인증서이다.

NoteBase

문 10. 네트워크 분석은 구조화를 통하여 개체 간 상호관계를 체계적으로 표현하고, 관계성을 파악하기 위한 분석 방법으로 정보학 분야에서 다양하게 활용되고 있다.
다음 중 네트워크에서 중심에 위치하는 정보를 표현하는 중심성 지표가 아닌 것은?

① 연결중심성
② 매개중심성
③ 근접중심성
④ 포괄중심성

정답 4번

출제 포인트

네트워크에서 노드의 중요성을 서로 다른 관점으로 측정하는 대표적인 중심성 지표를 구별하는 문제이다.

핵심 개념

중심성은 네트워크 안에서 특정 노드가 얼마나 중요한 위치를 차지하는지를 수치로 나타낸 지표이다.
연결중심성은 직접 연결된 이웃의 수, 매개중심성은 다른 노드 사이의 최단경로를 중개하는 정도, 근접중심성은 다른 모든 노드에 얼마나 가까운지를 측정한다.

선지별 해설

① O — 연결중심성은 한 노드에 직접 연결된 링크 또는 이웃 노드가 얼마나 많은지를 나타내는 대표적인 중심성 지표이다.
② O — 매개중심성은 한 노드가 다른 노드 쌍 사이의 최단경로에 얼마나 자주 포함되는지를 나타내므로 정보 흐름의 중개 역할을 측정한다.
③ O — 근접중심성은 한 노드에서 다른 노드들까지의 최단거리 합을 바탕으로 전체 네트워크에 얼마나 빠르게 도달할 수 있는지를 나타내는 중심성 지표이다.
④ X — 틀린 표현은 '포괄중심성'이다. 포괄중심성은 네트워크 분석의 대표적인 표준 중심성 지표가 아니다. 올바르게는 연결중심성·매개중심성·근접중심성이나 고유벡터중심성 등을 중심성 지표로 분류한다.

오답 포인트

④의 '포괄'이 여러 관계를 종합적으로 측정하는 전문용어처럼 보이는 점이 함정이다. 익숙한 일상어인지가 아니라 네트워크 분석에서 정의된 표준 지표명인지를 기준으로 판단한다.

다시 나오면 이렇게 판단

직접 연결 수는 연결중심성, 최단경로의 중개 역할은 매개중심성, 다른 노드까지의 거리 또는 도달 용이성은 근접중심성으로 연결한다. 이 세 가지 측정 기준에 해당하지 않는 낯선 명칭은 표준 중심성 개념인지 확인한다.

NoteBase

암기 포인트

연결은 이웃 수, 매개는 길목, 근접은 거리이다.

문 11. 데이터사이언스의 체계에 관한 다음 설명 중 가장 옳지 않은 것은?

① 데이터 관리영역과 기계학습영역으로 구분할 수 있다.

② 데이터사이언스 프로세스에는 사람이 개입하여 통제와 조절이 가능하다.

③ 인공지능의 구현은 지도학습 방식의 기계학습에 의존한다.

④ 대용량, 다양성, 빠른 속도의 특징을 가지는 빅데이터를 포함한다.

정답 3번

출제 포인트

데이터사이언스의 구성 영역과 처리 과정, 빅데이터의 특징을 이해하고 인공지능과 기계학습의 관계를 구분하는 문제이다.

핵심 개념

데이터사이언스는 데이터를 수집·저장·정제·관리하고, 분석과 기계학습을 통해 의미 있는 지식이나 예측을 도출하는 체계이다. 인공지능은 지능적 과업을 수행하는 시스템 전반을 뜻하며, 기계학습은 데이터를 이용해 성능을 개선하는 인공지능 구현 방법의 하나이다. 지도학습은 정답이 표시된 데이터로 학습하는 기계학습 방식이므로 인공지능 전체와 동일하지 않다.

선지별 해설

① O — 데이터사이언스는 데이터를 수집·저장·정제하는 데이터 관리영역과, 데이터에서 패턴을 학습하여 분류·예측 등에 활용하는 기계학습영역으로 구분하여 설명할 수 있다.
② O — 문제 정의, 데이터 선택과 전처리, 모델 선택·평가, 결과 해석과 적용 단계에서 사람이 판단하며 프로세스를 통제하고 조절할 수 있다.
③ X — 틀린 표현은 '지도학습 방식의 기계학습에 의존한다'이다. 인공지능은 지도학습만으로 구현되는 것이 아니며, 비지도학습·강화학습 등 다른 기계학습 방식과 규칙 기반 방법 등도 이용할 수 있다. 올바르게는 '지도학습은 인공지능을 구현하는 여러 방법 중 하나이다'라고 해야 한다.
④ O — 빅데이터는 일반적으로 대용량, 다양성, 빠른 생성·처리 속도라는 3V 특성을 가지며, 데이터사이언스가 수집·관리·분석하는 데이터의 범위에 포함된다.

오답 포인트

③은 '기계학습이 인공지능 구현에 활용된다'는 사실을 '인공지능은 지도학습에 의존한다'로 지나치게 좁힌 선지이다. 기계학습과 지도학습의 포함 관계를 뒤집지 않아야 한다.

NoteBase

다시 나오면 이렇게 판단

인공지능은 가장 넓은 범주, 기계학습은 그 구현 방법의 하나, 지도학습은 다시 기계학습의 한 방식으로 구분한다. ‘오직’, ‘의존한다’처럼 특정 방식만 가능하다고 단정하는 표현은 다른 구현 방식의 존재를 확인하여 판단한다.

암기 포인트

인공지능 ⊃ 기계학습 ⊃ 지도학습이며, 빅데이터의 기본 3V는 대용량·다양성·속도이다.

NoteBase

문 12. 다음 <보기>에서 메타 검색엔진에 대한 설명으로 옳은 것을 모두 고른 것은?

<보기>
ㄱ. 자체 검색기능 없이 여러 개의 검색엔진을 연계하여 이를 검색하여 보여준다.
ㄴ. 많은 양의 정보 검색이 가능하다.
ㄷ. 데이터의 중복도가 적다.
ㄹ. 검색 속도가 빠르다.

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄷ, ㄹ

정답 1번

출제 포인트

메타 검색엔진의 작동 방식과 그에 따른 장점·단점을 구별하는 문제이다.

핵심 개념

메타 검색엔진은 자체적으로 구축한 검색 데이터베이스를 이용하기보다 사용자의 질의를 여러 검색엔진에 보내고, 각 검색 결과를 모아 제시하는 검색 서비스이다.
ㄱ은 옳다. 여러 검색엔진을 연계해 결과를 보여 준다.
ㄴ은 옳다. 여러 검색엔진의 검색 범위를 함께 활용하므로 많은 양의 정보를 찾을 수 있다.
ㄷ은 옳지 않다. 서로 다른 검색엔진이 같은 자료를 반환할 수 있으므로 결과의 중복도가 높아질 수 있다.
ㄹ은 옳지 않다. 여러 검색엔진에 질의하고 결과를 수집·통합해야 하므로 일반적으로 검색 속도가 빠르다고 보기 어렵다.

선지별 해설

① O — ㄱ과 ㄴ이 모두 옳으므로 올바른 조합이다.
② X — 'ㄱ, ㄷ' 중 ㄷ이 틀렸다. 메타 검색 결과는 중복도가 적은 것이 아니라 여러 검색엔진의 결과가 겹쳐 중복도가 높아질 수 있으며, 올바른 조합은 ㄱ, ㄴ이다.
③ X — 'ㄴ, ㄷ' 중 ㄷ이 틀렸다. 많은 양의 정보 검색은 가능하지만 결과의 중복도는 높아질 수 있으며, 올바른 조합은 ㄱ, ㄴ이다.
④ X — 'ㄷ, ㄹ'은 모두 틀렸다. 메타 검색은 결과 중복이 생길 수 있고 여러 검색엔진의 결과를 모으는 과정 때문에 검색 속도가 빠르다고 보기 어려우며, 올바른 조합은 ㄱ, ㄴ이다.

NoteBase

오답 포인트

여러 검색엔진을 사용한다는 사실을 곧바로 '중복도가 적다' 또는 '속도가 빠르다'와 연결하면 틀린다. 검색 범위가 넓어지는 것은 장점이지만, 같은 결과가 겹치거나 결과 통합에 시간이 드는 것은 단점이다.

다시 나오면 이렇게 판단

'여러 검색엔진에 동시에 질의한다'는 구조에서 효과를 추론한다. 검색 범위와 정보량은 커지지만, 결과 중복과 통합 시간도 늘어날 수 있다고 판단한다.

암기 포인트

메타 검색은 여러 엔진을 한꺼번에 활용하므로 정보량은 많고, 중복 가능성과 처리 시간은 커진다.

NoteBase

문 13. 다음 <보기>에서 생성형 인공지능에 대한 설명으로 옳은 것을 모두 고른 것은?

<보기>
ㄱ. 머신러닝 알고리즘을 활용한다.
ㄴ. 학습 데이터를 기반으로 유사한 콘텐츠를 새롭게 만들어 내는 기술을 가리킨다.
ㄷ. 대형 언어모델이나 이미지 생성 모델을 활용한다.
ㄹ. 이미지 생성 모델로는 AE, GAN, BERT 등이 있다.

① ㄱ
② ㄱ, ㄴ
③ ㄱ, ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ, ㄹ

정답 3번

출제 포인트

생성형 인공지능의 정의와 대표 모델을 구분하고, 각 설명의 옳고 그름을 판별하는 문제이다.

핵심 개념

생성형 인공지능은 학습 데이터의 패턴과 구조를 머신러닝으로 익힌 뒤, 이를 바탕으로 텍스트·이미지 등 새로운 콘텐츠를 생성하는 인공지능이다. 대형 언어모델은 텍스트 생성에, 이미지 생성 모델은 이미지 생성에 활용된다.
ㄱ은 옳다. 생성형 인공지능은 머신러닝 알고리즘을 활용한다.
ㄴ은 옳다. 학습 데이터에서 익힌 패턴을 바탕으로 유사한 특성을 지닌 새로운 콘텐츠를 만든다.
ㄷ은 옳다. 생성형 인공지능에는 대형 언어모델과 이미지 생성 모델 등이 활용된다.
ㄹ은 틀리다. BERT는 주로 자연어의 문맥을 이해하도록 사전학습된 언어 모델이지 이미지 생성 모델이 아니다.
따라서 옳은 것은 ㄱ, ㄴ, ㄷ이다.

선지별 해설

① X — ㄱ은 옳지만, 옳은 설명인 ㄴ과 ㄷ을 빠뜨렸다. ‘ㄱ’이 아니라 ‘ㄱ, ㄴ, ㄷ’이 되어야 한다.
② X — ㄱ과 ㄴ은 옳지만, 대형 언어모델이나 이미지 생성 모델을 활용한다는 ㄷ도 옳다. ‘ㄱ, ㄴ’이 아니라 ‘ㄱ, ㄴ, ㄷ’이 되어야 한다.
③ O — ㄱ, ㄴ, ㄷ은 모두 생성형 인공지능의 원리와 활용 모델을 올바르게 설명하고, 틀린 ㄹ은 제외하였다.
④ X — ㄹ까지 옳다고 포함한 부분이 틀렸다. BERT는 이미지 생성 모델이 아니라 자연어 처리에 활용되는 사전학습 언어 모델이므로 ‘ㄱ, ㄴ, ㄷ, ㄹ’이 아니라 ‘ㄱ, ㄴ, ㄷ’이 되어야 한다.

NoteBase

오답 포인트

ㄹ은 AE·GAN과 함께 BERT를 나열하여 모두 이미지 생성 모델인 것처럼 보이게 한다. BERT의 핵심 용도가 자연어의 문맥 표현 학습이라는 점을 구분해야 한다. 또한 '옳은 것을 모두' 고르는 문제이므로 옳은 설명을 일부만 담은 ①과 ②도 정답이 아니다.

다시 나오면 이렇게 판단

먼저 각 보기의 진위를 개별적으로 판단한 뒤, 옳은 보기의 조합이 선택지와 정확히 일치하는지 확인한다. 모델 분류에서는 LLM·BERT처럼 언어를 다루는 모델과 GAN 등 이미지를 생성하는 데 쓰이는 모델의 주된 용도를 구분한다.

암기 포인트

생성형 AI는 학습한 패턴으로 새 콘텐츠를 만들며, BERT는 이미지 생성 모델이 아니라 언어 모델이다.

NoteBase

문 14. 다음 <보기>에 해당하는 기술은 무엇인가?

<보기>
- 거대언어모델에 쌓인 데이터와 별개 외부 데이터를 이용해 답변의 정확도를 높여주는 기술이다.
- 거대언어모델의 단점인 환각현상을 보완하는 역할을 한다.
- 이용자 질문인 프롬프트에 맞는 정확한 정보를 제공한다.

① 자연어 기반 코드 생성 모델
② 언어 생성 모델
③ 검색 증강 생성
④ 트랜스포머 모델

정답 3번

출제 포인트

외부 정보를 검색해 거대언어모델의 답변 생성에 활용하는 검색 증강 생성의 특징을 구별하는 문제이다.

핵심 개념

검색 증강 생성(RAG)은 이용자의 질문과 관련된 정보를 외부 데이터베이스나 문서에서 검색한 뒤, 그 정보를 거대언어모델의 답변 생성에 함께 활용하는 기술이다. 모델이 학습 당시 기억한 정보에만 의존하지 않으므로 답변의 정확성과 근거성을 높이고 환각현상을 줄이는 데 도움이 된다.

선지별 해설

① X — 자연어 기반 코드 생성 모델은 자연어로 작성한 요구사항을 바탕으로 프로그램 코드를 생성하는 모델이다. 외부 데이터를 검색해 답변 생성에 결합하는 기술이라는 표현으로 고치면 검색 증강 생성이다.
② X — 언어 생성 모델은 입력을 바탕으로 자연어 문장을 생성하는 모델을 넓게 가리킨다. 외부 정보를 먼저 검색하고 그 결과를 답변 생성에 활용하는 기술이라는 표현으로 고치면 검색 증강 생성이다.
③ O — 검색 증강 생성은 질문과 관련된 외부 데이터를 검색하여 거대언어모델의 생성 과정에 제공하므로 답변의 정확도를 높이고 환각현상을 보완한다.
④ X — 트랜스포머 모델은 어텐션을 중심으로 입력 요소 사이의 관계를 처리하는 신경망 구조이다. 외부 데이터를 검색해 생성 결과를 보강하는 기술이라는 표현으로 고치면 검색 증강 생성이다.

오답 포인트

보기에는 거대언어모델과 답변 생성이 언급되어 있어 ②나 ④로 혼동하기 쉽다. 그러나 정답을 가르는 표현은 '별개 외부 데이터 이용'과 '프롬프트에 맞는 정보 제공'이며, 이는 검색 결과를 생성에 결합하는 검색 증강 생성의 특징이다.

NoteBase

다시 나오면 이렇게 판단

문제에서 외부 문서 검색, 검색 결과의 생성 단계 결합, 최신·정확한 정보 제공, 환각현상 완화가 함께 제시되면 검색 증강 생성으로 판단한다.

암기 포인트

RAG는 '외부에서 검색하고, 근거를 붙여 생성한다'로 기억한다.

NoteBase

문 15. 과학기술정보, 연구데이터, 정보분석서비스 및 연구인프라를 연계·융합하여 연구자가 필요로 하는 지식인프라를 제공하는 한국과학기술정보연구원(KISTI)의 서비스는 무엇인가?

① ScienceON
② DataON
③ AccessON
④ NTIS

정답 1번

출제 포인트

과학기술 지식인프라 통합 서비스와 연구데이터·오픈액세스·국가연구개발정보 서비스를 구별하는 문제이다.

핵심 개념

ScienceON은 과학기술정보, 연구데이터, 정보분석서비스 및 연구인프라를 연계·융합하여 연구자에게 필요한 지식인프라를 제공하는 서비스이다.
DataON은 연구데이터의 검색·공유·활용을 지원하는 플랫폼이고, AccessON은 학술정보의 오픈액세스와 개방형 학술출판을 지원하는 플랫폼이며, NTIS는 국가연구개발사업 관련 정보를 통합 제공하는 서비스이다.

선지별 해설

① O — ScienceON은 문항에 제시된 여러 과학기술 자원과 서비스를 연계·융합해 연구자에게 지식인프라를 제공하므로 옳다.
② X — DataON은 문항의 통합 지식인프라 서비스명이 아니다. 틀린 표현은 'DataON'이며, 연구데이터의 검색·공유·활용을 지원하는 플랫폼을 가리킨다. 문항의 설명에 해당하는 것은 ScienceON이다.
③ X — AccessON은 문항의 통합 지식인프라 서비스명이 아니다. 틀린 표현은 'AccessON'이며, 학술정보의 오픈액세스와 개방형 학술출판을 지원하는 플랫폼을 가리킨다. 문항의 설명에 해당하는 것은 ScienceON이다.
④ X — NTIS는 문항의 통합 지식인프라 서비스명이 아니다. 틀린 표현은 'NTIS'이며, 국가연구개발사업 관련 정보를 통합 제공하는 서비스이다. 문항의 설명에 해당하는 것은 ScienceON이다.

오답 포인트

네 서비스 모두 과학기술·연구 정보를 다루어 혼동하기 쉽다. 문항의 핵심 표현은 여러 자원과 서비스를 '연계·융합'하여 종합적인 '지식인프라'를 제공한다는 점이며, 이는 ScienceON을 가리킨다.

다시 나오면 이렇게 판단

서비스가 다루는 핵심 대상을 먼저 찾는다. 통합 지식인프라는 ScienceON, 연구데이터는 DataON, 오픈액세스는 AccessON, 국가연구개발사업 정보는 NTIS로 구분한다.

NoteBase

암기 포인트

ScienceON은 과학기술 지식인프라 통합, DataON은 연구데이터, AccessON은 오픈액세스, NTIS는 국가연구개발사업 정보이다.

문 16. 문헌 클러스터링(Document Clustering)에 관한 다음 설명 중 가장 옳지 않은 것은?

① 문헌 클러스터링은 유사한 주제나 내용을 가진 문서들을 그룹화하는 기법이다.

② 문헌 클러스터링은 주로 텍스트 마이닝과 정보 검색에서 사용된다.

③ 문헌 클러스터링 기법에는 K-평균 클러스터링(K-means clustering)과 계층적 클러스터링(Hierarchical Clustering)이 포함된다.

④ 문헌 클러스터링은 클러스터의 수를 사전에 지정하지 않고 자동으로 결정된다.

정답 4번

출제 포인트

문헌 클러스터링의 목적과 활용 분야, 대표 기법, 클러스터 수의 결정 방식을 구분하는 문제이다.

핵심 개념

문헌 클러스터링은 문서의 내용이나 주제상 유사도를 기준으로 비슷한 문서끼리 묶는 비지도 학습 기법이다. 클러스터 수의 처리 방식은 알고리즘마다 다르다. K-평균은 일반적으로 클러스터 수 K를 미리 정해야 하고, 계층적 클러스터링은 계층 구조를 만든 뒤 절단 기준에 따라 클러스터 수를 정할 수 있다.

선지별 해설

① O — 문헌 클러스터링은 주제나 내용이 유사한 문서들을 같은 집단으로 묶는 기법이다.
② O — 대량의 문서에서 구조와 주제 집단을 발견하므로 텍스트 마이닝과 정보 검색에 활용된다.
③ O — K-평균 클러스터링과 계층적 클러스터링은 모두 문헌을 유사도에 따라 묶는 데 사용할 수 있는 대표적인 클러스터링 기법이다.
④ X — 틀린 표현은 '클러스터의 수를 사전에 지정하지 않고 자동으로 결정된다'이다. 모든 문헌 클러스터링 기법이 클러스터 수를 자동으로 결정하는 것은 아니며, K-평균은 일반적으로 원하는 클러스터 수 K를 사전에 지정해야 한다. 올바르게는 '클러스터 수의 결정 방식은 알고리즘에 따라 다르다'이다.

오답 포인트

클러스터링이 정답 라벨 없이 집단을 찾는 비지도 학습이라는 사실과 클러스터 수까지 항상 자동으로 정한다는 주장은 서로 다르다. ④는 '비지도 학습'을 '클러스터 수의 자동 결정'로 확대하여 표현한 것이 함정이다.

다시 나오면 이렇게 판단

클러스터 수에 관한 선지는 알고리즘별로 판단한다. K-평균은 일반적으로 K를 미리 정하고, 계층적 클러스터링은 생성된 계층 구조의 절단 기준에 따라 집단 수가 달라질 수 있으므로 '항상 자동 결정'과 같은 단정은 틀린 표현이다.

NoteBase

암기 포인트

문헌 클러스터링은 유사 문서 묶기이고, K-평균의 K는 일반적으로 사전 지정이다.

NoteBase

**문 17. 새로운 창작과 합법적인 정보의 공유를 촉진시키려는 목적으로 만들어진 크리에이티브 커먼즈 라이선스(Creative Commons License)가 인터넷에서 활용되고 있다.
다음 중 [저작자표시-변경금지]를 나타내는 크리에이티브 커먼즈 라이선스 유형을 올바르게 표시한 것은?**

① CC BY-NC

② CC BY-ND

③ CC BY-SA

④ CC BY-NC-SA

정답 2번

출제 포인트

크리에이티브 커먼즈 라이선스의 이용 조건을 나타내는 약어 BY, NC, ND, SA의 뜻을 구별하는 문제이다.

핵심 개념

BY는 저작자표시, NC는 비영리, ND는 변경금지, SA는 동일조건변경허락을 뜻한다. 따라서 [저작자표시-변경금지]는 BY와 ND를 결합한 CC BY-ND이다.

선지별 해설

① X — CC BY-NC는 저작자표시-비영리이다. 틀린 표현은 NC이며, 변경금지를 나타내려면 NC가 아니라 ND를 사용하여 CC BY-ND라고 해야 한다.
② O — CC BY-ND에서 BY는 저작자표시, ND는 변경금지를 뜻하므로 제시된 조건과 정확히 일치한다.
③ X — CC BY-SA는 저작자표시-동일조건변경허락이다. 틀린 표현은 SA이며, 변경금지를 나타내려면 SA가 아니라 ND를 사용하여 CC BY-ND라고 해야 한다.
④ X — CC BY-NC-SA는 저작자표시-비영리-동일조건변경허락이다. 틀린 표현은 NC-SA이며, 저작자표시-변경금지는 CC BY-ND라고 해야 한다.

오답 포인트

변경과 관련된 약어인 ND와 SA를 혼동하기 쉽다. ND는 변경 자체를 허용하지 않는 조건이고, SA는 변경한 저작물을 같은 라이선스 조건으로 공유하도록 하는 조건이다. 또한 NC는 변경 여부가 아니라 영리적 이용 여부를 제한한다.

다시 나오면 이렇게 판단

라이선스 이름의 조건을 순서대로 약어로 바꾼다. 저작자표시는 BY, 비영리는 NC, 변경금지는 ND, 동일조건변경허락은 SA로 대응시키면 된다.

NoteBase

암기 포인트

BY=저작자표시, NC=비영리, ND=변경금지, SA=동일조건변경허락이다.

NoteBase

문 18. 크롤러(Crawler)에 관한 다음 설명 중 가장 옳지 않은 것은?

① 크롤러는 주기적으로 검색된 정보를 데이터베이스화하는 역할을 담당한다.

② 크롤러는 웹상의 문서나 이미지, 영상 등을 수집하는 데 사용되는 검색 엔진 자체이다.

③ 크롤러는 정보를 수집하기 위해 자동화된 방식으로 제어된다.

④ 크롤러는 봇(Bot)이라고도 불리며, 웹상의 정보를 자동으로 검색하고 수집하는 프로그램이다.

정답 2번

출제 포인트

이 문제는 웹 정보를 자동으로 수집하는 크롤러와, 수집·색인·검색 결과 제공 기능을 포괄하는 검색 엔진을 구별하는 문제이다.

핵심 개념

크롤러는 웹 문서를 자동으로 찾아다니며 문서·이미지·영상 등의 정보를 수집하는 프로그램이다. 검색 엔진은 크롤러가 수집한 정보를 저장·색인하고 사용자의 질의에 맞는 결과를 제공하는 전체 시스템이다. 따라서 크롤러는 검색 엔진을 구성하는 요소이지 검색 엔진 자체가 아니다.

선지별 해설

① O — 크롤러는 웹을 주기적으로 방문하여 새 정보나 변경된 정보를 수집하고, 검색에 활용할 수 있도록 저장·색인되는 과정에 정보를 제공한다.
② X — 틀린 표현은 '검색 엔진 자체'이다. 크롤러는 웹상의 문서·이미지·영상 등을 수집하는 자동화 프로그램이며, 올바르게는 '검색 엔진을 구성하는 정보 수집 요소'라고 해야 한다.
③ O — 크롤러는 사람이 각 웹페이지를 직접 방문하는 방식이 아니라 미리 정한 규칙에 따라 링크를 따라가며 정보를 수집하도록 자동화되어 있다.
④ O — 크롤러는 웹 로봇 또는 봇이라고도 하며, 웹상의 정보를 자동으로 탐색하고 수집하는 프로그램이다.

오답 포인트

②는 크롤러가 검색 엔진에서 핵심적인 수집 기능을 수행한다는 점을 이용해 크롤러를 검색 엔진 전체와 동일시한 선지이다. 수집 프로그램과 검색 서비스를 제공하는 전체 시스템을 구분해야 한다.

다시 나오면 이렇게 판단

'웹을 돌아다니며 정보를 모은다'는 설명은 크롤러에 해당하고, '색인과 질의 처리를 거쳐 검색 결과를 제공한다'는 설명은 검색 엔진 전체에 해당한다고 판단한다.

NoteBase

암기 포인트

크롤러는 자동 수집 봇이고, 검색 엔진은 수집·색인·검색을 포괄하는 시스템이다.

NoteBase

문 19. 다음 중 IF(영향력 지수)를 계산하는 식으로 가장 올바른 것은?

① IF = 논문이 전체 인용된 횟수해당 저널에 발표된 전체 논문의 수

② IF = 해당 저널에 발표된 논문의 수논문이 전체 인용된 횟수

③ IF = 지난 2년간 해당 저널에 발표된 논문의 수지난 2년간 해당 저널에 발표된 논문의 인용 횟수

④ IF = 지난 2년간 해당 저널에 발표된 논문의 인용 횟수지난 2년간 해당 저널에 발표된 논문의 수

정답 4번

출제 포인트

영향력 지수(IF)의 산출 기간과 분자·분모를 정확히 구분하는 문제이다.

조건 정리

IF는 저널에 지난 2년간 발표된 논문이 받은 인용 횟수를 같은 기간에 그 저널이 발표한 논문 수로 나눈 값이다.
분자는 인용 횟수, 분모는 논문 수이며, 두 항목 모두 지난 2년이라는 범위를 반영한다.

사용할 공식

IF=지난 2년간 해당 저널에 발표된 논문의 인용 횟수지난 2년간 해당 저널에 발표된 논문의 수
IF는 논문 한 편당 평균적으로 얼마나 인용되었는지를 나타내므로 인용 횟수를 논문 수로 나눈다.

선지 판단

① X — 전체 인용 횟수와 전체 논문 수를 사용하여 지난 2년이라는 산출 범위를 반영하지 않았다. 틀린 표현인 '전체'는 '지난 2년간 해당 저널에 발표된'으로 고쳐야 한다.
② X — 논문 수를 분자에, 인용 횟수를 분모에 두어 분자와 분모가 뒤바뀌었고 지난 2년의 범위도 빠졌다.
올바르게는 지난 2년간의 인용 횟수를 분자에, 지난 2년간의 논문 수를 분모에 둔다.
③ X — 지난 2년의 범위는 반영했지만 논문 수를 인용 횟수로 나누어 분자와 분모를 거꾸로 배치했다.
올바르게는 지난 2년간의 인용 횟수를 지난 2년간의 논문 수로 나눈다.
④ O — 지난 2년간 해당 저널에 발표된 논문의 인용 횟수를 같은 기간에 발표된 논문 수로 나눈 식이므로 IF의 산식과 일치한다.

오답 포인트

③은 '지난 2년간'이라는 표현이 들어 있어 그럴듯하지만 분자와 분모가 뒤바뀐 식이다. ①은 분자와 분모의 방향은 맞아 보여도 산출 기간을 '전체'로 잡은 점이 함정이다.

NoteBase

다시 나오면 이렇게 판단

IF 산식은 먼저 기간이 지난 2년인지 확인하고, 다음으로 '논문 한 편당 인용 횟수'가 되도록 인용 횟수 ÷ 논문 수의 순서인지 확인한다.

암기 포인트

IF = 지난 2년 논문의 인용 횟수 ÷ 지난 2년 논문 수이다.

NoteBase

문 20. 다음 중 2차 정보원에 해당하지 않는 것은?

① 서지(Bibliography)

② 데이터베이스(Database)

③ 학술잡지(Scholarly Journal)

④ 색인(Index)

정답 3번

출제 포인트

이 문제는 원저자의 연구 성과를 직접 담은 1차 정보원과, 1차 정보원을 찾아 이용하도록 정리·가공한 2차 정보원을 구분하는 문제이다.

핵심 개념

1차 정보원은 연구자가 생산한 새로운 연구 결과나 원문을 직접 전달하는 자료이다. 학술논문을 수록하는 학술잡지가 대표적이다.
2차 정보원은 1차 정보원의 존재와 소재를 확인하고 검색할 수 있도록 서지사항 등을 조직·가공한 자료이다. 서지, 색인, 문헌 검색용 데이터베이스가 이에 해당한다.

선지별 해설

① X — 서지는 문헌의 서지사항을 일정한 기준에 따라 모아 배열하여 1차 정보원을 찾게 하는 2차 정보원이다. '2차 정보원에 해당하지 않는다'가 틀리며, 올바르게는 '서지는 2차 정보원에 해당한다'이다.
② X — 문헌 검색용 데이터베이스는 여러 1차 정보원의 서지정보 등을 조직하여 검색하게 하므로 2차 정보원이다. '2차 정보원에 해당하지 않는다'가 틀리며, 올바르게는 '데이터베이스는 2차 정보원에 해당한다'이다.
③ O — 학술잡지는 연구자의 새로운 연구 결과를 담은 학술논문을 직접 수록하는 대표적인 1차 정보원이므로 2차 정보원에 해당하지 않는다.
④ X — 색인은 주제어·저자명 등의 접근점을 제공하여 필요한 1차 정보원의 소재를 찾게 하는 2차 정보원이다. '2차 정보원에 해당하지 않는다'가 틀리며, 올바르게는 '색인은 2차 정보원에 해당한다'이다.

오답 포인트

데이터베이스에 원문이 들어 있을 수 있다는 점만 보고 모두 1차 정보원으로 판단하기 쉽다. 이 문항은 개별 수록 내용이 아니라 문헌 정보를 조직하고 검색하게 하는 정보원 유형을 묻고 있으므로 데이터베이스를 서지·색인과 함께 2차 정보원으로 구분한다.

NoteBase

다시 나오면 이렇게 판단

새로운 연구 내용이나 원문 자체를 직접 제공하면 1차 정보원, 그 원문을 찾도록 서지정보와 접근점을 정리해 제공하면 2차 정보원으로 판단한다.

암기 포인트

학술잡지는 연구 결과의 원문인 1차 정보원이고, 서지·데이터베이스·색인은 원문을 찾게 하는 2차 정보원이다.

NoteBase

문 21. 더빈(Dervin)의 의미형성(Sense-making)이론의 세 개의 핵심 개념이 바르게 짝지어진 것은?

① 상황 - 이용·도움 - 격차
② 상황 - 이용·도움 - 탐색
③ 정보 - 이용·도움 - 격차
④ 정보 - 이용·도움 - 탐색

정답 1번

출제 포인트

더빈(Dervin)의 의미형성이론을 구성하는 세 핵심 개념인 상황, 격차, 이용·도움을 구별하는 문제이다.

핵심 정리

더빈의 의미형성이론은 이용자가 특정 상황에서 지식의 불연속인 격차를 인식하고, 그 격차를 건너기 위해 정보를 이용하여 도움을 얻는 과정을 설명한다. 핵심 개념은 상황, 격차, 이용·도움이다.

선지별 해설

① O — 상황, 이용·도움, 격차는 더빈의 의미형성이론을 이루는 세 핵심 개념이므로 바르게 짝지어졌다.
② X — '탐색'이 잘못되었다. 올바르게는 상황, 이용·도움, 격차이다. 탐색은 정보행동의 한 과정일 수 있지만 이 모형의 세 핵심 개념 가운데 하나로 제시되는 말은 아니다.
③ X — '정보'가 잘못되었다. 올바르게는 정보가 아니라 '상황'이며, 전체 조합은 상황, 이용·도움, 격차이다.
④ X — '정보'와 '탐색'이 모두 잘못되었다. 올바르게는 각각 '상황'과 '격차'로 고쳐 상황, 이용·도움, 격차로 구성해야 한다.

출제자가 노린 함정

정보와 탐색은 정보행동론에서 자주 쓰이는 말이어서 핵심 개념처럼 보이지만, 더빈의 의미형성이론에서 고정적으로 연결해야 하는 세 용어는 상황, 격차, 이용·도움이다.

다시 나오면 이렇게 판단

더빈의 의미형성이론이 나오면 '상황에서 격차를 만나 정보를 이용해 도움을 얻는다'는 흐름으로 세 개념을 확인한다. 선지에 정보나 탐색이 핵심 개념으로 들어가면 상황 또는 격차가 빠졌는지 살핀다.

암기 포인트

더빈의 의미형성 = 상황·격차·이용(도움)이다.

NoteBase

문 22. 다음 중 더블린 코어 메타데이터 요소와 해당 요소에 대응하는 인코딩 스킴(Encoding Scheme)의 연결이 올바르지 않은 것은?

① date - W3C Date and Time Formats(W3CDTF)

② format - IMT

③ identifier - IP

④ subject - LCSH, MeSH

정답 3번

출제 포인트

더블린 코어 메타데이터 요소에 사용할 수 있는 인코딩 스킴을 정확히 연결하는 문제이다.

핵심 개념

인코딩 스킴은 메타데이터 값의 형식이나 통제어휘를 명확하게 지정하는 규칙이다. date에는 날짜·시간 표기 형식인 W3CDTF, format에는 인터넷 미디어 유형인 IMT, identifier에는 자원의 식별자를 나타내는 URI, subject에는 주제명 표목표인 LCSH와 MeSH 등을 연결할 수 있다.

선지별 해설

① O — date와 W3C Date and Time Formats(W3CDTF)의 연결은 옳다. W3CDTF는 날짜와 시간을 일정한 형식으로 표현하는 인코딩 스킴이다.
② O — format과 IMT의 연결은 옳다. IMT는 자원의 파일 형식이나 매체 유형을 인터넷 미디어 유형으로 표시할 때 사용한다.
③ X — identifier에 연결한 'IP'가 틀린 표현이다. 올바르게는 'identifier - URI'이다. URI는 자원을 식별하는 문자열의 형식을 지정한다.
④ O — subject와 LCSH, MeSH의 연결은 옳다. 두 스킴은 각각 일반 주제와 의학 분야의 통제된 주제명 표목을 제공한다.

오답 포인트

③의 IP와 URI를 모두 인터넷 자원과 관련된 용어로 보고 혼동하기 쉽다. IP는 네트워크상 장치의 주소 체계이고, 더블린 코어 identifier의 인코딩 스킴으로 연결되는 것은 URI이다.

다시 나오면 이렇게 판단

요소가 표현하는 값의 종류를 먼저 확인한다. 날짜는 W3CDTF, 매체 형식은 IMT, 자원 식별자는 URI, 주제명은 LCSH·MeSH로 연결한다.

NoteBase

암기 포인트

date-W3CDTF, format-IMT, identifier-URI, subject-LCSH·MeSH이다.

NoteBase

문 23. 오픈액세스 추진방법에 관한 다음 설명 중 가장 옳지 않은 것은?

① 골드 OA는 저자가 논문출판비용을 부담한다.

② 실버 OA는 저자가 학술논문을 디지털 저장소에 셀프 아카이빙한다.

③ 골드 OA는 출판 후 원문을 즉시 이용 할 수 있다.

④ 학술지 원문은 PDF 이외에 기계가독형 포맷으로 생산·공개되는 추세이다.

정답 2번

출제 포인트

오픈액세스의 골드 OA와 그린 OA를 공개 주체와 공개 방식에 따라 구별하는 문제이다.

핵심 개념

골드 OA는 학술지 출판사가 논문을 공개하여 독자가 출판된 원문을 즉시 이용하게 하는 방식이다. 대표적인 비용 모델에서는 저자 측이 논문처리비용을 부담한다.
그린 OA는 저자가 논문의 허용된 버전을 기관·주제별 디지털 저장소 등에 직접 기탁하는 셀프 아카이빙 방식이다.

선지별 해설

① O — 골드 OA의 대표적인 운영 모델에서는 독자에게 구독료를 받는 대신 저자 측이 논문처리비용을 부담한다.
② X — 틀린 표현은 '실버 OA'이다. 저자가 학술논문을 디지털 저장소에 직접 기탁하는 셀프 아카이빙은 그린 OA이다. 올바르게는 '그린 OA는 저자가 학술논문을 디지털 저장소에 셀프 아카이빙한다'이다.
③ O — 골드 OA는 출판사가 논문을 오픈액세스로 출판하므로 출판된 원문을 즉시 이용할 수 있다.
④ O — 학술지 원문은 화면 열람용 PDF뿐 아니라 검색·추출·재이용에 유리한 XML, HTML 등의 기계가독형 포맷으로도 생산·공개되는 추세이다.

오답 포인트

②는 '저자가 저장소에 직접 올린다'는 설명 자체는 맞지만 그 방식의 명칭을 실버 OA로 바꾸어 놓았다. 셀프 아카이빙이라는 단어가 나오면 그린 OA를 연결해야 한다.

다시 나오면 이렇게 판단

출판사가 OA 학술지에 공개하면 골드 OA, 저자가 저장소에 직접 기탁하면 그린 OA로 판단한다. 비용 부담, 공개 시점, 파일 포맷은 이 기본 구분과 별도로 확인한다.

암기 포인트

골드는 출판사 공개, 그린은 저장소 셀프 아카이빙이다.

NoteBase

문 24. 브라우징 검색에 관한 다음 설명 중 가장 옳지 않은 것은?

① 브라우징 검색은 주제가 명확히 정의되지 않았을 때 정보를 얻기 위해 활용한다.

② 브라우징 검색은 명확하게 설명하기 어려운 정보를 찾기 위해 활용한다.

③ 브라우징 검색은 적합정보와 부적합정보가 섞여있는 것으로부터 올바른 정보를 선택하기 위해 활용한다.

④ 브라우징 검색은 정보의 '우연한 발견(Serendipity)'이 가능하지 않는다.

정답 4번

출제 포인트

브라우징 검색을 사용하는 상황과 그 핵심 장점인 우연한 발견을 이해하는지를 묻는 문제이다.

핵심 개념

브라우징 검색은 검색어와 목표를 처음부터 정확히 정하기 어려울 때 정보 구조나 검색 결과를 훑어보며 관련 정보를 탐색하는 방식이다. 탐색 과정에서 처음에는 예상하지 못했던 유용한 정보를 발견하는 세렌디피티(Serendipity)가 가능하다는 점이 중요한 특징이다.

선지별 해설

① O — 주제가 명확히 정의되지 않았을 때 여러 정보와 범주를 살펴보며 검색 방향을 구체화할 수 있으므로 브라우징 검색을 활용할 수 있다.
② O — 필요한 정보를 정확한 검색어로 표현하기 어려운 경우에도 관련 항목을 훑고 연결을 따라가며 원하는 정보에 접근할 수 있다.
③ O — 적합정보와 부적합정보가 함께 제시된 결과를 살펴보면서 필요한 정보를 식별하고 선택하는 데 활용할 수 있다.
④ X — 틀린 표현은 '우연한 발견(Serendipity)이 가능하지 않는다'이다. 올바르게는 브라우징 검색 과정에서 예상하지 못한 유용한 정보를 우연히 발견할 수 있다.

오답 포인트

④는 '가능하지 않는다'라는 부정 표현으로 브라우징 검색의 대표적 장점을 정반대로 서술하였다. 문항도 '옳지 않은 것'을 묻고 있으므로 이중으로 부정 방향을 놓치기 쉽다.

다시 나오면 이렇게 판단

검색 목표나 검색어가 불명확하여 정보를 훑고 탐색해야 하는 상황, 그리고 탐색 중 예상 밖의 유용한 정보를 발견하는 상황은 브라우징 검색의 특징으로 판단한다.

NoteBase

암기 포인트

브라우징 검색은 '불명확한 요구를 탐색으로 구체화하고, 우연한 발견도 얻는 검색'이다.

NoteBase

문 25. 다음 표는 정보자원의 기술에 중점을 두는 전통적인 메타데이터와, 다양한 분야에서 생성되는 여러 가지 형태의 정보자원을 체계화된 구조를 통해 상세하게 기술하여 보다 종합적인 정보활동을 지원하는 차세대 메타데이터의 특징을 비교한 것이다. 빈 칸에 들어갈 내용으로 바르게 짝지어진 것은?

구 분	전통적 메타데이터	차세대 메타데이터
해석의 다양성	객관적	주관적
발생	정적	동적
요소의 비중	ㄱ	ㄴ
유연성	ㄷ	ㄹ
인코딩 기술	XML 기반	RDF 기반
콘텐츠 관리	중앙집중식	분산식

① ㄱ: 자원 중심적, ㄴ: 상황 중심적, ㄷ: 구문 의존적, ㄹ: 의미 의존적

② ㄱ: 상황 중심적, ㄴ: 자원 중심적, ㄷ: 구문 의존적, ㄹ: 의미 의존적

③ ㄱ: 자원 중심적, ㄴ: 상황 중심적, ㄷ: 의미 의존적, ㄹ: 구문 의존적

④ ㄱ: 상황 중심적, ㄴ: 자원 중심적, ㄷ: 의미 의존적, ㄹ: 구문 의존적

정답 1번

출제 포인트

전통적 메타데이터와 차세대 메타데이터를 요소의 비중과 유연성이라는 차이축으로 구분하는 문제이다.

핵심 개념

전통적 메타데이터는 개별 정보자원의 정해진 속성을 기술하는 자원 중심적 구조이며, XML 기반의 형식과 구조에 맞추는 구문 의존적 성격이 강하다.
차세대 메타데이터는 자원 사이의 관계와 이용 맥락까지 함께 표현하는 상황 중심적 구조이며, RDF를 통해 데이터의 뜻과 관계를 드러내는 의미 의존적 성격이 강하다.
따라서 ㄱ은 자원 중심적, ㄴ은 상황 중심적, ㄷ은 구문 의존적, ㄹ은 의미 의존적이다.

NoteBase

선지별 해설

① O — 전통적 메타데이터의 자원 중심성과 구문 의존성, 차세대 메타데이터의 상황 중심성과 의미 의존성을 모두 바르게 연결하였다.
② X — ㄱ과 ㄴ을 서로 바꾼 것이 틀렸다. 올바르게는 ㄱ이 자원 중심적이고 ㄴ이 상황 중심적이다.
③ X — ㄷ과 ㄹ을 서로 바꾼 것이 틀렸다. 올바르게는 ㄷ이 구문 의존적이고 ㄹ이 의미 의존적이다.
④ X — 요소의 비중과 유연성을 모두 반대로 연결하였다. 올바르게는 ㄱ 자원 중심적, ㄴ 상황 중심적, ㄷ 구문 의존적, ㄹ 의미 의존적이다.

오답 포인트

XML과 RDF가 모두 구조화된 표현 기술이라는 점만 보고 ㄷ과 ㄹ을 혼동하기 쉽다. XML 기반의 전통적 메타데이터는 정해진 형식과 문법인 구문에 더 의존하고, RDF 기반의 차세대 메타데이터는 자원의 뜻과 관계인 의미에 더 무게를 둔다.

다시 나오면 이렇게 판단

표의 다른 대비를 함께 묶어 판단한다. 전통적 메타데이터는 객관적·정적·자원 중심적·구문 의존적·중앙집중식이고, 차세대 메타데이터는 주관적·동적·상황 중심적·의미 의존적·분산식이다.

암기 포인트

전통은 자원·구문·XML, 차세대는 상황·의미·RDF로 연결된다.

NoteBase