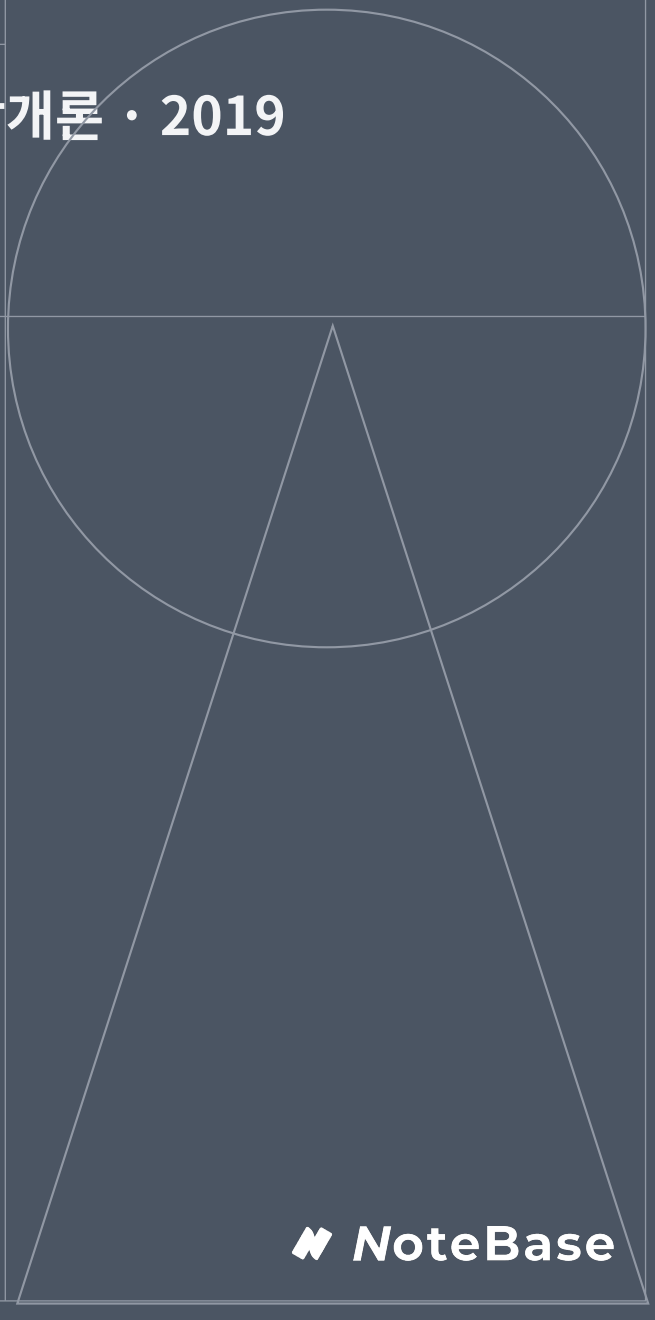
기출OX 해설집

공무원/국회직/9급 · 정보학개론 · 2019



NoteBase

기출OX 해설집

공무원/국회직/9급 · 정보학개론 · 2019

공무원/국회직/9급 · 정보학개론 · 2019

1. 다음 설명에 부합하는 용어는?

> 2012년 ISO 27729 표준으로 지정된 식별자로, 작가, 연구자, 실연자, 영상제작자 등 모든 창작자의 공개 신분을 식별하기 위한 국제표준 기호이다. 총 16자리 숫자문자열로 구성되고, 우리나라의 경우 국립중앙도서관이 2016년부터 발급을 시작하여, 영화진흥위원회, 한국복제전송저작권협회, 한국연구재단, 한국과학기술정보연구원, 한국교육학술정보원 등의 기관 API 데이터와 연계를 확대하고 있다.

① ISNI(International Standard Name Identifier)
② ISCI(International Standard Creator Identifier)
③ BIC(Biographical Identification Code)
④ IAI(International Author Identifier)
⑤ ISAN(International Standard Author Number)

정답 1번

출제 포인트

ISO 27729에 따라 창작자의 공개 신원을 식별하는 16자리 국제표준 식별기호의 명칭과 약어를 묻는 문제이다.

핵심 개념

ISNI는 International Standard Name Identifier의 약어이다. 작가·연구자·실연자·영상제작자 등 창작자의 공개 신원을 구별하기 위한 16자리 국제표준 식별기호이며, 제시문에서는 ISO 27729가 결정적 단서이다.

NoteBase

선지별 해설

① O - ISNI는 International Standard Name Identifier의 약어로, 제시문의 ISO 27729 및 창작자 공개 신원 식별이라는 설명과 일치한다.
② X - 틀린 표현은 'ISCI(International Standard Creator Identifier)'이다. 제시된 표준의 올바른 명칭은 'ISNI(International Standard Name Identifier)'이다.
③ X - 틀린 표현은 'BIC(Biographical Identification Code)'이다. 제시문이 말하는 16자리 국제표준 식별기호는 'ISNI(International Standard Name Identifier)'이다.
④ X - 틀린 표현은 'IAI(International Author Identifier)'이다. 대상이 작가뿐 아니라 연구자·실연자·영상제작자 등으로 넓으며, 올바른 용어는 'ISNI(International Standard Name Identifier)'이다.
⑤ X - 틀린 표현은 'ISAN(International Standard Author Number)'이다. 제시문의 ISO 27729에 해당하는 용어는 'ISNI(International Standard Name Identifier)'이다.

오답 포인트

②·④·⑤는 Creator, Author처럼 제시문의 대상과 비슷한 단어를 넣어 그럴듯하게 만든 선지이다. 그러나 표준명은 대상의 직업명이 아니라 공개 신원을 뜻하는 Name을 사용한 International Standard Name Identifier이다.

다시 나오면 이렇게 판단

ISO 27729, 16자리, 창작자·연구자 등의 공개 신원 식별이라는 단서가 함께 나오면 ISNI를 선택한다. 약어와 영문 전체 명칭이 모두 International Standard Name Identifier로 정확히 대응하는지도 확인한다.

암기 포인트

ISO 27729의 공개 신원(Name) 식별자 = ISNI.

NoteBase

2. A저널은 2012년에 창간되었다. 창간 이후 A저널에 게재된 논문의 수와 2018년 기준 A저널을 인용한 논문의 수는 아래의 표와 같다. 이 표를 근거로 할 때 2018년 기준 A저널의 피인용반감기(Cited Half-Life)는?

연도	A저널 게재 논문 수	2018년 기준 A저널을 인용한 논문 수
2012	70	80
2013	80	90
2014	100	50
2015	120	40
2016	90	40
2017	110	30
2018	130	10
계	700	340

① 3년
② 4년
③ 5년
④ 6년
⑤ 7년

정답 3번

출제 포인트

이 문제는 특정 연도에 받은 전체 인용의 절반에 이를 때까지 최신 발행 연도부터 과거로 거슬러 올라가며 피인용 논문 수를 누적하여 피인용반감기를 구하는 문제이다.

조건 정리

주어진 값: 2018년 기준 A저널을 인용한 논문은 총 340편이다.
구해야 할 값: 최신 연도인 2018년부터 누적한 피인용 논문 수가 전체의 절반에 도달하는 데 포함되는 연수이다.
주의할 조건: A저널의 연도별 게재 논문 수가 아니라 '2018년 기준 A저널을 인용한 논문 수'를 누적한다.

사용할 공식

전체 인용의 절반은 다음과 같다.
$340 \div 2 = 170$
피인용반감기는 최신 연도부터 과거로 거슬러 올라가며 누적 피인용 수가 170편에 도달할 때까지의 연수를 세는 이유는, 전체 인용의 절반이 어느 기간에 집중되어 있는지를 나타내는 지표이기 때문이다.

NoteBase

계산 과정

1. 2018년부터 역순으로 누적하면 10편이다.
2. 2017년까지 누적하면 10+30=40편, 2016년까지는 80편, 2015년까지는 120편이다.
3. 2014년의 50편을 더하면 170편이 되어 전체 340편의 절반에 도달한다.
4. 2014년부터 2018년까지는 2014·2015·2016·2017·2018의 5개 연도이므로 피인용반감기는 5년이다.

선지 판단

① X — 3년 동안인 2016~2018년의 누적 피인용 수는 40+30+10=80편으로, 절반인 170편에 이르지 못한다. '3년'이 아니라 '5년'이 옳다.
② X — 4년 동안인 2015~2018년의 누적 피인용 수는 40+40+30+10=120편으로, 170편에 이르지 못한다. '4년'이 아니라 '5년'이 옳다.
③ O — 2014~2018년의 누적 피인용 수는 50+40+40+30+10=170편으로 전체 340편의 정확히 절반이며, 이 기간은 5년이다.
④ X — 누적 피인용 수는 2014년까지 거슬러 올라간 5년째에 이미 170편에 도달하므로 6년까지 포함할 필요가 없다. '6년'이 아니라 '5년'이 옳다.
⑤ X — 누적 피인용 수는 5년째에 절반에 도달하므로 창간 이후의 7개 연도를 모두 포함하지 않는다. '7년'이 아니라 '5년'이 옳다.

오답 포인트

표의 'A저널 게재 논문 수' 합계 700을 사용하거나, 2012년부터 정방향으로 누적하면 틀린다. 또한 2014년과 2018년의 차이만 계산하여 4년으로 고르기 쉽지만, 양 끝 연도를 모두 포함하므로 2014~2018년은 5년이다.

다시 나오면 이렇게 판단

피인용반감기는 피인용 수 합계의 50%를 먼저 구한 뒤, 기준연도부터 과거 방향으로 피인용 수를 누적하고, 그 값에 처음 도달하는 연도까지 양 끝 연도를 포함하여 센다.

암기 포인트

피인용반감기 = 전체 피인용 수의 절반을 최신 연도부터 역누적하여 채우는 기간이다.

NoteBase

3. 다음 <보기>에서 디지털 아카이빙 방법과 설명이 옳게 짝지어진 것을 모두 고른 것은?

<보 기>

ㄱ. 복제(replication) - 자료의 백업
ㄴ. 마이그레이션(migration) - 원자료의 형식과 구조 변화 가능
ㄷ. 인캡슐레이션(encapsulation) - 새로운 저장매체 이용
ㄹ. 에뮬레이션(emulation) - 저장매체의 물리적 약화에 대비

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄱ, ㄹ
④ ㄴ, ㄷ
⑤ ㄴ, ㄹ

정답 1번

출제 포인트

디지털 아카이빙의 보존 방법인 복제, 마이그레이션, 인캡슐레이션, 에뮬레이션의 목적과 특징을 구별하는 문제이다.

핵심 개념

복제(replication)는 디지털 자료의 사본을 만들어 원본의 손실에 대비하는 방법이며, 핵심은 백업이다.
마이그레이션(migration)은 자료를 새로운 시스템이나 형식으로 옮기면서 필요에 따라 파일 형식과 구조를 변환하는 방법이다.
인캡슐레이션(encapsulation)은 디지털 객체를 해석하는 데 필요한 메타데이터, 소프트웨어 정보 등을 객체와 함께 묶어 보존하는 방법이다.
에뮬레이션(emulation)은 과거의 하드웨어·소프트웨어 환경을 현재 시스템에서 모방하여 원래 자료를 이용할 수 있게 하는 방법이다.

NoteBase

선지별 해설

① O — ㄱ의 복제는 자료의 사본을 만드는 백업 방식이고, ㄴ의 마이그레이션은 자료를 옮기는 과정에서 형식과 구조를 변환할 수 있으므로 둘 다 옳다.
② X — ㄱ은 옳지만 ㄷ의 '새로운 저장매체 이용'은 인캡슐레이션의 설명이 아니다. 올바르게는 인캡슐레이션은 디지털 객체와 그 해석에 필요한 메타데이터·환경 정보를 함께 묶어 보존하는 방법이다.
③ X — ㄱ은 옳지만 ㄹ의 '저장매체의 물리적 약화에 대비'는 에뮬레이션의 설명이 아니다. 올바르게는 에뮬레이션은 구형 하드웨어·소프트웨어 환경을 모방하여 기술적 노후화에 대응하는 방법이다.
④ X — ㄴ은 옳지만 ㄷ이 틀리므로 옳은 설명을 모두 고른 조합이 아니다. ㄷ은 '새로운 저장매체 이용'이 아니라 '디지털 객체와 메타데이터·환경 정보를 함께 묶어 보존'으로 고쳐야 한다.
⑤ X — ㄴ은 옳지만 ㄹ이 틀리므로 옳은 설명을 모두 고른 조합이 아니다. ㄹ은 '저장매체의 물리적 약화에 대비'가 아니라 '과거의 실행 환경을 모방하여 기술적 노후화에 대응'으로 고쳐야 한다.

오답 포인트

ㄷ은 저장매체를 새것으로 바꾸는 행위와 인캡슐레이션을 혼동하게 하고, ㄹ은 저장매체의 물리적 열화와 실행 환경의 기술적 노후화를 혼동하게 한다. 인캡슐레이션의 핵심은 관련 정보의 '묶음'이고, 에뮬레이션의 핵심은 과거 환경의 '모방'이다.

다시 나오면 이렇게 판단

사본 생성은 복제, 형식·구조 변환은 마이그레이션, 객체와 환경 정보의 묶음은 인캡슐레이션, 구형 실행 환경의 모방은 에뮬레이션으로 구별한다.

암기 포인트

복제는 복사, 마이그레이션은 변환·이동, 인캡슐레이션은 묶음, 에뮬레이션은 환경 모방이다.

NoteBase

4. 디지털 도서관에서 저작권 관리를 위해 제품마다 서로 다른 유일한 코드를 부여하여 불법으로 유통되는 복사본들을 감시하거나 추적하는 방법은?

① Watermarking
② Fingerprinting
③ Fair Use
④ Creative Commons License
⑤ IPLeft

정답 2번

출제 포인트

이 문제는 디지털 저작권 관리 방법 가운데 복사본이나 이용자마다 서로 다른 식별 코드를 넣어 불법 유통의 출처를 추적하는 기술을 구별하는 문제이다.

핵심 개념

Fingerprinting은 배포되는 복사본이나 이용자마다 고유한 식별정보를 삽입하여 불법 복제물이 발견되었을 때 유출 경로를 추적하는 기술이다.
Watermarking은 저작권자나 소유권 등의 정보를 콘텐츠에 눈에 잘 띄지 않게 삽입하여 권리 표시·소유권 입증·무단 이용 억제에 활용하는 기술이다.
Fair Use는 일정한 요건 아래 저작권자의 허락 없이도 저작물을 이용할 수 있도록 하는 공정이용 원칙이다.
Creative Commons License는 저작자가 이용 허용 조건을 미리 표시하는 표준화된 공개 라이선스이다.
IPLeft는 지식과 정보의 독점에 비판적 입장에서 정보공유와 대안적 지식재산권 제도를 추구하는 단체·운동이다.

NoteBase

선지별 해설

① X — Watermarking은 저작권자 등의 정보를 콘텐츠에 삽입하여 권리를 표시하거나 소유권을 입증하는 데 초점이 있다. 틀린 표현은 'Watermarking'이며, 복사본마다 서로 다른 고유 코드를 넣어 유출 경로를 추적하는 기술은 'Fingerprinting'이다.
② O — Fingerprinting은 복사본이나 이용자별로 고유한 식별정보를 부여하므로 불법 유통된 복제물에서 배포·유출 경로를 찾아낼 수 있다.
③ X — Fair Use는 일정한 요건을 충족하면 허락 없이 저작물을 이용할 수 있게 하는 저작권 제한 원칙이지, 복사본별 추적 코드를 부여하는 기술이 아니다. 틀린 표현은 'Fair Use'이며, 올바르게는 'Fingerprinting'이다.
④ X — Creative Commons License는 저작자가 저작물의 이용 조건을 미리 제시하는 공개 라이선스이지, 불법 복제물을 감시·추적하는 식별 기술이 아니다. 틀린 표현은 'Creative Commons License'이며, 올바르게는 'Fingerprinting'이다.
⑤ X — IPLeft는 정보공유와 대안적 지식재산권 제도를 추구하는 단체·운동이지, 제품별 고유 코드를 삽입하는 기술이 아니다. 틀린 표현은 'IPLeft'이며, 올바르게는 'Fingerprinting'이다.

오답 포인트

① Watermarking과 ② Fingerprinting은 모두 디지털 콘텐츠에 정보를 삽입할 수 있어 혼동하기 쉽다. 그러나 문제의 '제품마다 서로 다른 유일한 코드'와 '추적'이라는 표현은 복사본·이용자별 식별정보를 사용하는 Fingerprinting을 가리킨다.

다시 나오면 이렇게 판단

저작권자 표시나 소유권 입증이 중심이면 Watermarking, 복사본·구매자별 고유 식별정보로 유출자를 찾아내는 것이 중심이면 Fingerprinting으로 판단한다.

암기 포인트

워터마킹은 권리 표시, 핑거프린팅은 복사본별 식별과 유출 추적이다.

NoteBase

5. 다음 <보기>는 데이터 모델의 관점에서 XML(eXtensible Markup Language)과 RDF(Resource Description Framework)를 비교한 설명이다. <보기>에서 옳은 것을 모두 고른 것은?

<보 기>

ㄱ. XML은 순서를 중요시하는 트리 구조이지만, RDF는 주어부, 술어부, 목적부로 이루어지는 트리플 구조의 객체 간 관계지향 모델이다.

ㄴ. XML에서는 태그의 배치 순서가 다르면 서로 다른 문서로 인식되지만, RDF에서는 각각의 트리플이 독립적으로 존재하므로 배치 순서는 중요하지 않다.

ㄷ. XML 스키마는 문서의 의미적 해석이 주된 기능이지만, RDF 스키마는 주로 구문적 해석에 사용된다.

ㄹ. XML은 트리 구조로 되어 있어 검색하기 복잡하지만, RDF는 독립적인 트리플 집합이므로 XML보다 검색이 용이하다.

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

정답 3번

출제 포인트

XML과 RDF를 데이터 구조, 순서의 의미, 스키마의 역할, 검색 방식이라는 기준으로 비교하는 문제이다.

핵심 개념

XML은 요소와 속성을 계층적으로 배치하는 순서 있는 트리 기반 문서 모델이다. 자식 요소의 배치 순서는 문서 구조의 일부가 된다.

RDF는 자원 사이의 관계를 주어-술어-목적어 트리플로 나타내는 그래프 기반 데이터 모델이다. RDF 그래프에서 트리플을 직렬화한 순서는 의미에 영향을 주지 않는다.

XML 스키마는 XML 문서에 허용되는 요소, 속성, 자료형, 출현 순서 등 구조와 문법을 정의한다. RDF 스키마는 클래스와 속성 및 이들의 관계를 정의하여 RDF 어휘에 의미를 부여한다. 따라서 ㄱ, ㄴ, ㄹ은 옳고 ㄷ은 두 스키마의 역할을 서로 바꾸어 서술했으므로 틀리다.

NoteBase

선지별 해설

① X
ㄱ과 ㄴ은 옳지만, RDF가 관계를 트리플 집합으로 표현하여 구조적 관계 검색에 유리하다는 옳은 설명 ㄹ이 빠져 있다. 틀린 부분은 'ㄱ, ㄴ'만 고른 것이며, 올바른 조합은 'ㄱ, ㄴ, ㄹ'이다.

② X
ㄱ은 옳지만 ㄷ은 XML 스키마와 RDF 스키마의 역할을 뒤바꾼 설명이다. 또한 옳은 ㄴ과 ㄹ이 빠져 있다. 틀린 조합 'ㄱ, ㄷ'은 'ㄱ, ㄴ, ㄹ'로 고쳐야 한다.

③ O
ㄱ은 XML의 순서 있는 트리 구조와 RDF의 주어-술어-목적어 관계 구조를 바르게 비교한다. ㄴ은 XML의 요소 배치 순서와 달리 RDF 그래프에서는 트리플의 나열 순서가 의미를 바꾸지 않는다는 점을 바르게 설명한다. ㄹ은 계층 경로 중심의 XML과 비교할 때 RDF가 독립적인 트리플을 연결해 관계 중심으로 질의하기 쉽다는 점을 바르게 설명한다. 따라서 옳은 것은 ㄱ, ㄴ, ㄹ이다.

④ X
ㄱ과 ㄹ은 옳지만 ㄷ이 틀렸고, 옳은 ㄴ이 빠져 있다. 틀린 부분은 'ㄷ'을 포함한 것이며, XML 스키마는 구조적·구문적 제약을, RDF 스키마는 의미적 관계를 정의한다고 고쳐야 한다. 올바른 조합은 'ㄱ, ㄴ, ㄹ'이다.

⑤ X
ㄴ과 ㄹ은 옳지만 ㄷ이 틀렸고, 옳은 ㄱ이 빠져 있다. 틀린 부분은 'ㄷ'을 포함한 것이며, 두 스키마의 역할을 바로잡고 ㄱ을 포함하면 올바른 조합은 'ㄱ, ㄴ, ㄹ'이 된다.

오답 포인트

핵심 함정은 ㄷ의 '의미적 해석'과 '구문적 해석'을 서로 바꿔 놓은 데 있다. XML 스키마는 문서가 어떤 구조와 형식을 따라야 하는지를 정하고, RDF 스키마는 클래스와 속성의 의미 관계를 정한다. 또한 RDF를 XML 문법으로 직렬화할 수 있다는 사실과 RDF 데이터 모델 자체가 순서 없는 트리플 그래프라는 사실을 혼동하기 쉽다.

다시 나오면 이렇게 판단

XML은 '순서 있는 계층 문서와 구조 검증', RDF는 '순서 없는 트리플 그래프와 관계·의미 표현'이라는 두 축으로 판단한다. 스키마가 나오면 XML 스키마는 문서 구조, RDF 스키마는 클래스·속성의 의미 관계에 연결한다.

암기 포인트

XML은 순서 있는 트리와 구조 스키마, RDF는 순서 없는 트리플 그래프와 의미 스키마이다.

NoteBase

6. 다음 <보기>에서 디지털 도서관과 관련된 표준 중 이용 통계와 관련된 것을 모두 고른 것은?

<보 기>

ㄱ. SRU	ㄴ. COUNTER	ㄷ. ODRL
ㄹ. LDAP	ㅁ. SUSHI	

① ㄱ, ㄷ
② ㄱ, ㄹ
③ ㄴ, ㄹ
④ ㄴ, ㅁ
⑤ ㄷ, ㅁ

정답 4번

출제 포인트

디지털 도서관 관련 표준 가운데 이용 통계의 작성·제공과 자동 수집에 직접 관련된 표준을 구별하는 문제이다.

핵심 개념

COUNTER: 전자정보자원의 이용 통계를 일관된 방식으로 기록하고 보고하기 위한 기준이다.
SUSHI: 이용 통계 보고서, 특히 COUNTER 보고서를 시스템 간에 자동으로 요청하고 수집하기 위한 프로토콜이다.
SRU: 웹 환경에서 검색 질의를 보내고 검색 결과를 받기 위한 정보검색 프로토콜이다.
ODRL: 디지털 콘텐츠의 허용, 금지, 의무 등 권리·정책 정보를 표현하기 위한 언어이다.
LDAP: 네트워크의 디렉터리 서비스에 접근하여 이용자나 조직 등의 정보를 조회·관리하기 위한 프로토콜이다.
따라서 이용 통계의 보고 기준은 COUNTER이고, 그 통계를 자동으로 수집하는 방식은 SUSHI이다.

NoteBase

선지별 해설

① X
ㄱ SRU와 ㄷ ODRL은 이용 통계 표준이 아니다. SRU는 정보검색, ODRL은 권리·정책 표현에 관련된다. ‘ㄱ, ㄷ’은 ‘ㄴ, ㅁ’으로 고쳐야 한다.

② X
ㄱ SRU는 이용 통계가 아니라 검색·검색 결과 전송에 관한 프로토콜이다. ㄹ LDAP도 이용 통계가 아니라 디렉터리 서비스 접근에 관한 프로토콜이다. ‘ㄱ, ㄹ’은 ‘ㄴ, ㅁ’으로 고쳐야 한다.

③ X
ㄴ COUNTER는 이용 통계와 관련되지만, ㄹ LDAP는 디렉터리 정보의 조회·관리에 관련되므로 조합이 틀리다. ‘ㄴ, ㄹ’은 ‘ㄴ, ㅁ’으로 고쳐야 한다.

④ O
ㄴ COUNTER는 전자정보자원 이용 통계 보고의 표준화 기준이고, ㅁ SUSHI는 그러한 이용 통계 보고서를 자동으로 요청·수집하는 프로토콜이다. 두 항목 모두 이용 통계와 직접 관련된다.

⑤ X
ㅁ SUSHI는 이용 통계와 관련되지만, ㄷ ODRL은 디지털 콘텐츠의 권리와 이용 정책을 표현하는 언어이다. ‘ㄷ, ㅁ’은 ‘ㄴ, ㅁ’으로 고쳐야 한다.

오답 포인트

디지털 도서관에서 쓰이는 표준이라는 공통점만 보고 모두 이용 통계 표준으로 묶기 쉽다. 이 문항에서는 각 약어의 기능을 구분해야 하며, 특히 COUNTER는 통계 보고의 기준이고 SUSHI는 그 보고서를 자동 수집하는 전달 방식이라는 짝이 핵심이다.

다시 나오면 이렇게 판단

‘이용량을 어떤 형식과 기준으로 보고하는가’이면 COUNTER, ‘그 이용 통계 보고서를 시스템이 자동으로 가져오는가’이면 SUSHI로 판단한다. 검색은 SRU, 권리 표현은 ODRL, 디렉터리 접근은 LDAP로 분리한다.

암기 포인트

COUNTER는 이용 통계를 세고 보고하는 기준, SUSHI는 그 통계 보고서를 자동으로 가져오는 프로토콜이다.

NoteBase

7. 다음은 각 문헌에 출현한 색인어를 나타낸다. 문헌 2가 검색되는 질의식으로 옳은 것은?

문헌 1: 범주화, 운영, 정보활용, 학교도서관
문헌 2: 교수매체, 정보매체, 정보활용, 학교도서관
문헌 3: 운영, 정보검색, 학교시설, 활용수업
문헌 4: 범주화, 이론, 정보검색, 정보활용

① 정보검색 OR 운영
② 정보검색 AND 정보활용
③ 정보활용 AND NOT 학교도서관
④ 정보검색 OR (학교도서관 AND 운영)
⑤ 학교도서관 AND NOT 범주화

정답 5번

출제 포인트

불리언 검색식의 OR, AND, NOT을 적용해 각 식의 검색 결과 집합을 구하고, 문헌 2만 남기는 식을 찾는 문제이다.

핵심 개념

OR는 두 검색어 중 하나라도 포함한 문헌을 찾는 합집합 연산이다. AND는 두 검색어를 모두 포함한 문헌을 찾는 교집합 연산이다. AND NOT은 앞 검색어를 포함한 문헌에서 뒤 검색어를 포함한 문헌을 제외하는 차집합 연산이다. 따라서 '학교도서관 AND NOT 범주화'는 학교도서관이 있으면서 범주화는 없는 문헌을 뜻한다.

선지별 해설

① X — '정보검색 OR 운영'은 정보검색이 있는 문헌 3·4와 운영이 있는 문헌 1·3을 합쳐 문헌 1·3·4를 검색하므로 문헌 2만 검색하지 못한다. 틀린 부분은 이 검색식 전체이며, 문헌 2만 남기려면 '학교도서관 AND NOT 범주화'로 고쳐야 한다.
② X — '정보검색 AND 정보활용'은 두 색인어를 모두 가진 문헌 4만 검색한다. 따라서 문헌 2를 검색하지 못하며, 문헌 2만 남기는 올바른 식은 '학교도서관 AND NOT 범주화'이다.
③ X — '정보활용 AND NOT 학교도서관'은 정보활용이 있는 문헌 1·2·4에서 학교도서관이 있는 문헌 1·2를 제외하여 문헌 4만 검색한다. 따라서 문헌 2만 남기는 올바른 식은 '학교도서관 AND NOT 범주화'이다.
④ X — '정보검색 OR (학교도서관 AND 운영)'에서 정보검색은 문헌 3·4를, 학교도서관과 운영의 교집합은 문헌 1을 검색하므로 최종 결과는 문헌 1·3·4이다. 문헌 2만 남기는 올바른 식은 '학교도서관 AND NOT 범주화'이다.
⑤ O — 학교도서관이 있는 문헌은 1·2이고, 이 가운데 범주화가 있는 문헌 1을 제외하면 문헌 2만 남는다.

NoteBase

오답 포인트

문헌 2에 '정보활용'과 '학교도서관'이 있다는 사실만 보고 이 검색어들을 단순히 AND로 묶으면 다른 문헌이 남거나 문헌 2가 제외될 수 있다. 목표 문헌에 있으면서 비교 문헌에도 있는 검색어를 먼저 잡고, 목표 문헌에는 없지만 비교 문헌에는 있는 검색어를 NOT으로 제외해야 한다.

다시 나오면 이렇게 판단

각 검색어의 문헌 집합을 적은 뒤 OR는 합집합, AND는 교집합, AND NOT은 차집합으로 계산한다. 특정 문헌 하나만 찾아야 하면 그 문헌을 포함하는 공통 검색어로 후보를 모으고, 원하지 않는 문헌만 가진 검색어를 NOT 조건으로 뺀다.

암기 포인트

OR는 더하고, AND는 겹치고, NOT은 뺀다. 학교도서관 {1, 2}에서 범주화 {1, 4}를 빼면 {2}이다.

NoteBase

8. 다음은 메타데이터에 대한 설명이다. ㉠, ㉡에 들어갈 용어로 옳은 것은?

> (㉠)은/는 메타데이터 요소와 특정 목적을 위해 정의된 사용규칙의 집합을 말한다. 즉, (㉠)은/는 정보자원을 기술하기 위해 마련된 일련의 기술규칙으로 이루어지며, 이 규칙은 메타데이터를 구현할 수 있는 구조를 정의한다.
>
> (㉡)은/는 메타데이터 요소들을 어떻게 기계 가독 형식으로 인코딩해야 하는가를 말한다. 즉, (㉡)은/는 공통의 포맷을 제공하여, 로컬 시스템에 저장된 데이터를 다른 시스템과 상호 교환하기 위해 이용된다.

	㉠	㉡
①	메타데이터 스키마	의미 구조(semantics)
②	메타데이터 스키마	전송 규칙(transfer rule)
③	메타데이터 스키마	구문 구조(syntax)
④	메타데이터 스킴	의미 구조(semantics)
⑤	메타데이터 스킴	내용 규칙(content rule)

정답 3번

출제 포인트

메타데이터 스키마와 구문 구조의 정의를 구별하여 두 빈칸에 알맞은 용어를 연결하는 문제이다.

핵심 개념

메타데이터 스키마: 메타데이터 요소와 그 사용 규칙을 체계화하여 메타데이터를 구현할 구조를 정의한 것이다. 키워드는 '요소·사용 규칙·구조'이다.
구문 구조(syntax): 메타데이터 요소를 기계가 판독할 수 있도록 인코딩하는 형식이다. 서로 다른 시스템이 데이터를 교환할 수 있게 하는 공통 표현 형식이라는 점이 핵심이다.
의미 구조(semantics)는 메타데이터 요소가 무엇을 뜻하는지를 다루며, 전송 규칙은 데이터 전송 방식을, 내용 규칙은 요소에 기록할 값의 형식과 내용을 다룬다. 따라서 이 문항의 인코딩 형식과는 다르다.

NoteBase

선지별 해설

① X — (ㄱ)의 '메타데이터 스키마'는 맞지만, (ㄴ)의 '의미 구조(semantics)'가 틀렸다. 의미 구조는 요소의 뜻을 나타낸다. 기계 판독이 가능한 공통 인코딩 형식은 '구문 구조(syntax)'이다.
② X — (ㄱ)은 맞지만, (ㄴ)의 '전송 규칙(transfer rule)'이 틀렸다. 전송 규칙은 데이터 전송 방식을 가리키며, 제시문의 공통 인코딩 형식은 '구문 구조(syntax)'이다.
③ O — 메타데이터 요소와 사용 규칙의 집합은 '메타데이터 스키마'이고, 요소의 기계 판독형 인코딩 형식은 '구문 구조(syntax)'이므로 두 설명에 모두 부합한다.
④ X — (ㄱ)의 '메타데이터 스킴'과 (ㄴ)의 '의미 구조(semantics)'가 모두 틀렸다. 올바른 연결은 '메타데이터 스키마—구문 구조(syntax)'이다.
⑤ X — (ㄱ)의 '메타데이터 스킴'과 (ㄴ)의 '내용 규칙(content rule)'이 모두 틀렸다. 내용 규칙은 요소에 입력할 값의 형식이나 내용을 정하는 규칙이다. 올바른 연결은 '메타데이터 스키마—구문 구조(syntax)'이다.

오답 포인트

'공통의 포맷'이라는 표현만 보고 전송 규칙을 고르거나, 메타데이터의 뜻과 관련 있다고 생각하여 의미 구조를 고르기 쉽다. 그러나 제시문의 결정적 표현은 '기계 판독 형식으로 인코딩'이며, 이는 구문 구조(syntax)를 뜻한다. 또한 '스키마'와 선택지의 '스킴'을 혼동하지 않아야 한다.

다시 나오면 이렇게 판단

'요소와 사용 규칙의 집합·구조 정의'가 나오면 스키마로 판단하고, '기계 판독·인코딩·공통 포맷·시스템 간 교환'이 나오면 구문 구조로 판단한다. '뜻'은 의미 구조, '입력할 값의 형식과 내용'은 내용 규칙으로 구별한다.

암기 포인트

스키마는 설계 구조, syntax는 인코딩 형식이다.

NoteBase

9. 헤팅 츄(Heting Chu)가 제시한 정보 표현과 검색 발전의 네 가지 주요 단계에 대한 설명으로 옳지 않은 것은?

① '급속한 성장' 단계는 컴퓨터의 등장, 온라인 시스템의 출현 등 전산화된 정보검색을 가능하게 한 시기이다.
② '정보요구의 증가' 단계는 기술보고서와 문헌 양의 급격한 증가와 밀접한 관련이 있다.
③ '일반화 단계'는 최종 이용자가 사서의 도움 없이 온라인 시스템에서 스스로 탐색하는 것을 가능하게 한 시기이다.
④ '네트워크 시대'는 인터넷으로 상징되며 네트워크로 연결된 정보검색을 가능하게 한 시기이다.
⑤ 시기적으로 '급속한 성장', '정보요구의 증가', '일반화 단계', '네트워크 시대' 순서로 발전하였다.

정답 5번

출제 포인트

이 문제는 헤팅 츄가 구분한 정보 표현과 검색 발전의 네 단계가 각각 어떤 특징을 가지며 어떤 순서로 전개되었는지를 묻는다.

핵심 개념

정보요구의 증가 단계는 과학기술 활동의 확대에 따라 기술보고서와 문헌이 급증하면서 정보를 체계적으로 찾아야 할 필요가 커진 시기이다.
급속한 성장 단계는 컴퓨터와 온라인 시스템의 등장으로 전산화된 정보검색이 가능해진 시기이다.
일반화 단계는 전문가의 중개에 주로 의존하던 검색이 최종 이용자의 직접 온라인 탐색으로 확대된 시기이다.
네트워크 시대는 인터넷과 정보 네트워크를 통해 서로 연결된 정보자원을 검색하는 시기이다.
발전 순서는 정보요구의 증가 → 급속한 성장 → 일반화 → 네트워크 시대이다.

선지별 해설

① O — '급속한 성장' 단계에는 컴퓨터와 온라인 시스템이 등장하여 전산화된 정보검색이 가능해졌으므로 옳다.
② O — '정보요구의 증가' 단계는 과학기술의 발전으로 기술보고서와 문헌의 양이 급격히 늘어나 정보검색의 필요성이 커진 것과 밀접하므로 옳다.
③ O — '일반화' 단계에는 최종 이용자가 사서와 같은 정보전문가의 중개 없이 온라인 시스템에서 직접 탐색할 수 있게 되었으므로 옳다.
④ O — '네트워크 시대'는 인터넷의 확산과 정보 네트워크의 연결을 바탕으로 네트워크형 정보검색이 가능해진 시기이므로 옳다.
⑤ X — 틀린 부분은 '급속한 성장'이 '정보요구의 증가'보다 먼저라는 순서이다. 올바른 순서는 '정보요구의 증가 → 급속한 성장 → 일반화 → 네트워크 시대'이다.

NoteBase

오답 포인트

①과 ②의 설명 자체는 모두 맞지만 ⑤가 두 단계의 시간적 순서를 바꾸어 제시했다. 문헌 증가로 검색의 필요성이 먼저 커지고, 그 요구에 대응하는 컴퓨터·온라인 검색 기술이 뒤이어 급속히 성장했다는 인과관계를 구분해야 한다.

다시 나오면 이렇게 판단

각 단계의 대표 표지를 연결하여 판단한다. 문헌 증가와 검색 필요성은 정보요구의 증가, 컴퓨터·온라인 시스템은 급속한 성장, 최종 이용자의 직접 탐색은 일반화, 인터넷과 연결망은 네트워크 시대에 해당한다. 순서 문제에서는 필요의 발생이 기술적 성장보다 먼저라는 흐름을 확인한다.

암기 포인트

문헌 증가로 요구가 생기고, 컴퓨터로 검색이 성장하며, 이용자에게 일반화된 뒤, 인터넷 네트워크로 연결된다.

NoteBase

10. 다음은 문헌 클러스터링의 종류에 대한 설명이다. ㉠, ㉡에 들어갈 용어로 옳은 것은?

> 문헌의 자동분류에서 배타적 분류는 (㉠) 분류와 (㉡) 분류로 구분되는데, (㉠) 분류는 분류 대상물에 대한 데이터 행렬이나 대상물 간의 유사도 행렬만을 사용하는 비지도 학습 분류에 해당하며, (㉡) 분류는 유사도 행렬 이외에 사전에 부여된 범주명을 함께 사용하는 지도 학습 분류를 의미한다.

	㉠	㉡
①	계층적	비계층적
②	내재적	외재적
③	절대적	상대적
④	배제적	상호적
⑤	정적	동적

정답 2번

출제 포인트

문헌 자동분류에서 데이터 행렬이나 유사도만 사용하는 내재적 분류와 사전에 부여된 범주명을 함께 사용하는 외재적 분류를 구별하는 문제이다.

핵심 개념

내재적 분류: 문헌의 데이터 행렬이나 문헌 간 유사도 행렬만으로 군집을 형성하는 비지도 학습 분류이다.
키워드는 '유사도만 사용'과 '비지도 학습'이다.
외재적 분류: 문헌의 유사도 정보에 더하여 사전에 부여된 범주명을 학습하여 분류하는 지도 학습 분류이다.
키워드는 '범주명 사용'과 '지도 학습'이다.
내재적 분류는 정답 범주를 미리 주지 않고 자료 내부의 구조를 찾으며, 외재적 분류는 미리 주어진 범주명을 학습에 이용한다.

NoteBase

선지별 해설

① X — '계층적·비계층적'은 군집을 단계적 구조로 만드는지에 따른 군집화 방식의 구분이므로, 비지도 학습과 지도 학습을 대비한 빈칸의 개념이 아니다. 틀린 표현 '계층적·비계층적'은 '내재적·외재적'으로 고쳐야 한다.
② O — 내재적 분류는 데이터 행렬이나 유사도 행렬만 사용하는 비지도 학습이고, 외재적 분류는 사전에 부여된 범주명을 함께 사용하는 지도 학습이므로 두 빈칸에 정확히 대응한다.
③ X — '절대적·상대적'은 이 문항에서 설명한 문헌 자동분류의 두 유형 명칭이 아니다. 틀린 표현 '절대적·상대적'은 '내재적·외재적'으로 고쳐야 한다.
④ X — '배제적·상호적'은 유사도 정보와 범주명 사용 여부를 기준으로 한 문헌 자동분류의 두 유형이 아니다. 틀린 표현 '배제적·상호적'은 '내재적·외재적'으로 고쳐야 한다.
⑤ X — '정적·동적'은 시간에 따른 변화 여부를 나타내는 일반적 대비로, 비지도 학습과 지도 학습의 구분을 설명하지 못한다. 틀린 표현 '정적·동적'은 '내재적·외재적'으로 고쳐야 한다.

오답 포인트

①의 '계층적·비계층적'도 실제 군집화 관련 용어이므로 혼동하기 쉽다. 그러나 지문은 군집의 구조가 아니라 사전에 부여된 범주명을 학습에 사용하는지를 기준으로 구분한다.

다시 나오면 이렇게 판단

범주명 없이 데이터와 유사도만으로 묶으면 내재적 분류이고, 미리 부여된 범주명을 학습에 함께 사용하면 외재적 분류로 판단한다.

암기 포인트

내재적은 자료 내부의 유사도만, 외재적은 외부에서 부여한 범주명까지 사용한다.

NoteBase

11. 질의나 문헌에서 의미적 정보를 추출하여 검색하는 방식으로, 사용자의 검색 질의에 나타난 의도와 키워드의 의미 분석을 통해 검색결과를 제시하는 방식은?

① 데이터 마이닝
② 클러스터링
③ 토픽 모델링
④ 시맨틱 검색
⑤ 메타 검색

정답 4번

출제 포인트

검색 질의의 의도와 키워드의 문맥적 의미를 분석해 검색 결과를 제시하는 방식을 구별하는 문제이다.

핵심 개념

시맨틱 검색은 단순한 문자열 일치를 넘어 질의의 의도, 단어의 의미, 문맥과 개념 사이의 관계를 파악하여 관련성 높은 결과를 찾는 검색 방식이다.
데이터 마이닝은 대규모 데이터에서 유용한 패턴이나 지식을 발견하는 활동이고, 클러스터링은 유사한 대상을 같은 집단으로 묶는 기법이다. 토픽 모델링은 문서 집합에 숨어 있는 주제 구조를 추정하며, 메타 검색은 여러 검색 엔진의 결과를 모아 통합해 보여 준다.

선지별 해설

① X — '데이터 마이닝'이라는 명칭이 틀렸다. 데이터 마이닝은 대규모 데이터에서 규칙, 관계, 패턴 등의 유용한 지식을 추출하는 활동이다. 제시된 설명의 올바른 용어는 '시맨틱 검색'이다.
② X — '클러스터링'이라는 명칭이 틀렸다. 클러스터링은 정답표 없이 특성이 유사한 데이터나 문서를 여러 집단으로 묶는 기법이다. 제시된 설명의 올바른 용어는 '시맨틱 검색'이다.
③ X — '토픽 모델링'이라는 명칭이 틀렸다. 토픽 모델링은 여러 문서에 함께 나타나는 단어의 분포를 바탕으로 잠재된 주제를 찾아내는 기법이다. 제시된 설명의 올바른 용어는 '시맨틱 검색'이다.
④ O — 시맨틱 검색은 사용자의 질의에 담긴 의도와 키워드의 의미 및 문맥을 분석하여 의미상 관련된 검색 결과를 제시하므로 문제의 설명과 일치한다.
⑤ X — '메타 검색'이라는 명칭이 틀렸다. 메타 검색은 하나의 질의를 여러 검색 엔진에 전달하고 각 검색 결과를 수집·통합하여 제시하는 방식이다. 제시된 설명의 올바른 용어는 '시맨틱 검색'이다.

NoteBase

오답 포인트

‘의미적 정보를 추출한다’는 표현 때문에 데이터 마이닝이나 토픽 모델링을 떠올리기 쉽다. 그러나 이 문항의 결정적 표현은 ‘사용자의 검색 질의에 나타난 의도’와 ‘키워드의 의미 분석을 통해 검색 결과를 제시’한다는 부분이며, 이는 시맨틱 검색의 특징이다.

다시 나오면 이렇게 판단

질문의 의도·문맥·의미를 이해해 결과를 찾으면 시맨틱 검색, 데이터에서 패턴을 찾으면 데이터 마이닝, 비슷한 대상을 묶으면 클러스터링, 문서의 잠재 주제를 찾으면 토픽 모델링, 여러 검색 엔진의 결과를 합치면 메타 검색으로 판단한다.

암기 포인트

의미와 의도를 이해하는 검색은 시맨틱 검색이다.

NoteBase

12. 다음의 도서관 네트워크 구성요소를 제시한 사람은?

- 참여도서관이 납부하는 봉사료에 의해 유지
- 전임직원의 지시 및 감독
- 독자적인 집행기구에 의해 운영
- 기계가독형 서지 데이터베이스를 상호협동으로 유지·증보
- 온라인 시스템으로 직·간접적으로 연결

① 보그만(Christine L. Borgman)
② 카츠(William A. Katz)
③ 사라세빅(Tefko Saracevic)
④ 스티븐스(Norman D. Stevens)
⑤ 셰라(Jesse H. Shera)

정답 4번

출제 포인트

이 문제는 도서관 네트워크의 운영·재정·기술적 구성요소를 제시한 학자를 연결하는 문제이다.

핵심 정리

노먼 D. 스티븐스: 참여 도서관의 분담금, 전임 직원의 지시·감독, 독자적 집행기구, 상호협동으로 유지·증보되는 기계가독형 서지 데이터베이스, 온라인 연결을 도서관 네트워크의 구성요소로 제시하였다.
크리스틴 L. 보그만: 디지털 도서관과 학술 커뮤니케이션, 정보 이용 연구와 관련된 학자이다.
윌리엄 A. 카츠: 참고봉사와 참고면담 분야를 대표하는 학자이다.
테프코 사라세빅: 정보검색의 적합성 및 정보과학의 성격과 관련된 연구를 전개하였다.
제시 H. 셰라: 사회인식론을 바탕으로 도서관학의 사회적 지식 조직과 이용을 설명하였다.

NoteBase

선지별 해설

① X — 보그만은 디지털 도서관과 학술 커뮤니케이션 연구로 알려진 학자이며, 제시문과 같은 도서관 네트워크 구성요소를 제시한 인물이 아니다. 틀린 연결은 '보그만'이며, 올바르게는 '노먼 D. 스티븐스'이다.
② X — 카츠는 참고봉사와 참고면담 분야의 대표 학자이므로 제시문의 네트워크 구성요소와 연결되지 않는다. 틀린 연결은 '카츠'이며, 올바르게는 '노먼 D. 스티븐스'이다.
③ X — 사라세빅은 정보검색의 적합성과 정보과학 연구를 대표하므로 제시문의 네트워크 운영 모형을 제시한 인물이 아니다. 틀린 연결은 '사라세빅'이며, 올바르게는 '노먼 D. 스티븐스'이다.
④ O — 스티븐스는 참여관의 분담금, 전임 직원, 독립 집행기구, 협동 서지 데이터베이스, 온라인 연결을 도서관 네트워크의 구성요소로 제시하였다.
⑤ X — 셰라는 사회인식론을 통해 도서관학의 사회적 지식 조직을 설명한 학자이므로 제시문의 네트워크 구성요소와 연결되지 않는다. 틀린 연결은 '셰라'이며, 올바르게는 '노먼 D. 스티븐스'이다.

출제자가 노린 함정

보그만·사라세빅·셰라처럼 정보학에서 자주 등장하는 학자들을 함께 제시하여 연구 영역을 혼동하게 한 것이 함정이다. 특히 '기계가독형 서지 데이터베이스'와 '온라인 연결'이라는 표현만 보고 디지털 도서관 연구자인 보그만을 고르면 안 된다.

다시 나오면 이렇게 판단

분담금으로 유지되는 참여관 협력체, 전임 직원과 독립 집행기구, 공동 서지 데이터베이스, 온라인 연결이 한 묶음으로 제시되면 스티븐스의 도서관 네트워크 구성요소로 판단한다.

암기 포인트

스티븐스 = 분담금·전임 직원·독립 집행기구·공동 서지 DB·온라인 연결이다.

NoteBase

13. OSI 참조 모델 7계층에서 데이터가 최적의 경로로 전송될 수 있도록 경로를 배정하고 구간별 혼잡을 제어하는 기능을 수행하는 계층은?

① 네트워크 계층
② 데이터링크 계층
③ 세션 계층
④ 전송 계층
⑤ 표현 계층

정답 1번

출제 포인트

OSI 7계층에서 경로 배정(라우팅)과 구간별 혼잡 제어를 담당하는 계층을 구별하는 문제이다.

핵심 개념

네트워크 계층은 패킷이 목적지까지 갈 경로를 선택하고 전달하는 계층이며, 라우팅과 혼잡 제어가 핵심 기능이다.
데이터링크 계층은 인접한 노드 사이의 프레임 전송을, 전송 계층은 종단 간 데이터 전송을 담당한다.
세션 계층은 통신 세션의 설정·유지·종료와 동기화를, 표현 계층은 데이터 형식 변환·암호화·압축을 담당한다.

선지별 해설

① O — 네트워크 계층은 목적지까지의 최적 경로를 결정하는 라우팅을 수행하고 패킷 혼잡을 제어하므로 제시된 기능에 해당한다.
② X — 틀린 표현은 '데이터링크 계층'이다. 데이터링크 계층은 인접 노드 간 프레임 전송, 물리 주소 지정, 오류·흐름 제어를 담당하며, 제시된 기능은 네트워크 계층의 기능이다.
③ X — 틀린 표현은 '세션 계층'이다. 세션 계층은 응용 프로그램 사이의 세션을 설정·유지·종료하고 대화 동기화를 담당하며, 경로 배정과 혼잡 제어는 네트워크 계층이 담당한다.
④ X — 틀린 표현은 '전송 계층'이다. 전송 계층은 송신지와 수신지 사이의 종단 간 전송, 분할·재조립, 오류·흐름 제어를 담당하며, 패킷의 경로 배정은 네트워크 계층이 담당한다.
⑤ X — 틀린 표현은 '표현 계층'이다. 표현 계층은 데이터 형식 변환, 암호화·복호화, 압축·해제를 담당하며, 경로 배정과 혼잡 제어는 네트워크 계층이 담당한다.

오답 포인트

④ 전송 계층도 전송 제어와 관련되어 있어 혼동하기 쉽다. 그러나 문제의 결정적 표현인 '최적의 경로로 전송될 수 있도록 경로를 배정'은 라우팅을 뜻하므로 네트워크 계층으로 판단한다.

NoteBase

다시 나오면 이렇게 판단

‘경로 선택·라우팅·패킷·논리 주소·혼잡 제어’가 나오면 네트워크 계층, ‘종단 간 신뢰성·분할·재조립’이 나오면 전송 계층으로 구분한다.

암기 포인트

경로를 고르는 라우팅은 네트워크 계층이다.

NoteBase

14. 다음 <보기>에서 정보검색 이론 및 기술의 발전에 공헌한 인물과 업적이 옳게 짝지어진 것만을 모두 고른 것은?

<보 기>

ㄱ. Hans P. Luhn - KWIC 색인 기법 제안
ㄴ. Calvin N. Mooers - 조합색인 기법 창안
ㄷ. Gerard Salton - SMART 시스템 개발
ㄹ. Eugene Garfield - ISI 설립

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

정답 3번

출제 포인트

정보검색 이론과 기술의 발전에 기여한 인물을 대표 업적과 정확히 연결할 수 있는지를 묻는 문제이다.

핵심 정리

ㄱ. Hans P. Luhn: KWIC 색인을 제안하였다. KWIC는 문헌의 핵심어를 그 앞뒤 문맥과 함께 제시하는 색인 방식이다.
ㄴ. Calvin N. Mooers: 정보검색이라는 용어와 Zatocoding으로 대표되는 인물이다. 조합색인 기법의 창안자는 Mortimer Taube이므로 제시된 연결은 틀리다.
ㄷ. Gerard Salton: 자동 정보검색 실험 시스템인 SMART 시스템을 개발하였다.
ㄹ. Eugene Garfield: 인용색인을 발전시키고 Institute for Scientific Information(ISI)을 설립하였다.
따라서 옳게 연결된 것은 ㄱ, ㄷ, ㄹ이다.

NoteBase

선지별 해설

① X — ㄴ을 포함한 것이 틀리다. ㄴ의 '조합색인 기법 창안'은 Calvin N. Mooers가 아니라 Mortimer Taube와 연결해야 하며, 올바른 전체 조합은 ㄱ, ㄷ, ㄹ이다.
② X — ㄴ을 포함하고 ㄷ을 제외한 것이 틀리다. Gerard Salton의 SMART 시스템 개발은 옳고, 올바른 전체 조합은 ㄱ, ㄷ, ㄹ이다.
③ O — Hans P. Luhn의 KWIC 색인, Gerard Salton의 SMART 시스템, Eugene Garfield의 ISI 설립이 모두 옳게 연결되어 있다.
④ X — ㄴ을 포함하고 ㄱ을 제외한 것이 틀리다. Hans P. Luhn의 KWIC 색인 제안은 옳으며, 올바른 전체 조합은 ㄱ, ㄷ, ㄹ이다.
⑤ X — 네 항목을 모두 옳다고 본 것이 틀리다. ㄴ의 'Calvin N. Mooers-조합색인 기법 창안'을 'Mortimer Taube-조합색인 기법 창안'으로 고쳐야 하므로, 올바른 전체 조합은 ㄱ, ㄷ, ㄹ이다.

출제자가 노린 함정

Calvin N. Mooers도 초기 정보검색 기술에 기여했기 때문에 ㄴ이 그럴듯해 보인다. 그러나 Mooers의 대표 연결어는 정보검색과 Zatocoding이고, 조합색인은 Mortimer Taube와 연결된다는 구분이 핵심이다.

다시 나오면 이렇게 판단

인물별 대표 연결어를 먼저 대조한다. Luhn-KWIC, Mooers-Zatocoding·정보검색, Taube-조합색인, Salton-SMART, Garfield-인용색인·ISI로 대응시키면 조합을 빠르게 판별할 수 있다.

암기 포인트

Luhn은 KWIC, Taube는 조합색인, Salton은 SMART, Garfield는 ISI이다.

NoteBase

15. 정보검색 모델의 특징에 대한 설명 중 나머지 넷과 다른 정보검색 모델을 설명하는 것은?

① 용어 가중치로 질의와 문헌을 표현한 뒤 유사도를 계산하여 결과물을 검색한다.
② 용어들의 집합관계를 나타낼 수 없다.
③ 불리언 논리를 이해하거나 사용할 필요가 없는 것이 장점이다.
④ 질의나 문헌을 분해하여 용어의 조작을 지원함으로써 검색의 유연성을 달성할 수 있다.
⑤ 가중치 부여 메커니즘이 주관적이거나 복잡해질 수 있다.

정답 4번

출제 포인트

벡터공간모형과 불리언모형의 검색 방식 및 장단점을 구별하는 문제이다.

핵심 개념

벡터공간모형은 질의와 문헌을 용어 가중치로 이루어진 벡터로 표현하고, 두 벡터의 유사도를 계산하여 검색 결과의 순위를 정하는 모형이다. 불리언모형은 용어를 AND, OR, NOT 등의 불리언 연산자로 결합하여 조건에 맞는 문헌을 검색하는 모형이다. 벡터공간모형의 핵심은 가중치·유사도·순위화이고, 불리언모형의 핵심은 용어의 논리적 결합과 조작이다.

NoteBase

선지별 해설

① X
질의와 문헌을 용어 가중치 벡터로 표현하고 유사도를 계산하여 결과를 검색하는 것은 벡터공간모형의 대표적 특징이다. 따라서 나머지 넷과 다른 모형을 설명한 선지가 아니다.

② X
용어를 서로 독립된 차원으로 취급하므로 용어 사이의 집합관계나 의미관계를 직접 나타내기 어렵다는 것은 벡터공간모형의 한계이다. 따라서 나머지 넷과 다른 모형을 설명한 선지가 아니다.

③ X
벡터공간모형은 질의를 가중치 벡터로 표현하므로 이용자가 AND, OR, NOT과 같은 불리언 논리를 직접 이해하거나 사용할 필요가 적다. 따라서 나머지 넷과 다른 모형을 설명한 선지가 아니다.

④ O
질의나 문헌을 용어 단위로 나누고 용어를 논리 연산자로 결합·조작하여 검색 조건을 구성하는 것은 불리언모형의 특징이다. 벡터공간모형의 유연성은 이러한 용어의 논리적 조작이 아니라 용어 가중치와 유사도에 따른 순위화에서 나온다.

⑤ X
용어 가중치의 산정 방식에 따라 검색 결과가 달라질 수 있고 가중치 부여가 복잡해질 수 있다는 것은 벡터공간모형의 단점이다. 따라서 나머지 넷과 다른 모형을 설명한 선지가 아니다.

오답 포인트

④의 '검색의 유연성'만 보면 벡터공간모형으로 오인하기 쉽다. 그러나 결정적인 표현은 '용어의 조작'이다. 논리 연산자를 이용한 용어 조작은 불리언모형에 해당하고, 가중치와 유사도 계산은 벡터공간모형에 해당한다.

다시 나오면 이렇게 판단

선지에서 '가중치·벡터·유사도·순위화'가 보이면 벡터공간모형으로, 'AND·OR·NOT·집합·용어의 논리적 조작'이 보이면 불리언모형으로 판단한다.

암기 포인트

벡터공간모형은 가중치와 유사도, 불리언모형은 논리 연산과 용어 조작이다.

NoteBase

16. 다음 <보기>에서 멀티미디어 검색을 위한 내용기반 색인에 사용될 수 있는 자료 형태와 속성(자질)이 옳게 짝지어진 것을 모두 고른 것은?

<보 기>

ㄱ. 이미지 - 대상, 크기, 이미지 유형
ㄴ. 음악 - 작곡가, 멜로디, 연주악기
ㄷ. 오디오 - 높낮이, 크기, 발음
ㄹ. 비디오 - 색분포, 질감, 돌출점

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄱ, ㄹ
④ ㄴ, ㄷ
⑤ ㄷ, ㄹ

정답 5번

출제 포인트

멀티미디어 검색을 위한 내용기반 색인에서 자료의 실제 내용으로부터 추출하는 자질과 제목·작곡가·파일 유형 같은 기술 정보를 구별하는 문제이다.

핵심 개념

내용기반 색인은 멀티미디어 자체의 신호나 시각적 특징을 분석하여 검색용 자질을 만드는 방식이다.
ㄱ X: 이미지의 대표적인 내용 자질은 색상, 질감, 형태 등이다. '이미지 유형'은 파일이나 형식을 설명하는 정보이므로 제시된 묶음 전체를 내용 자질로 볼 수 없다. 올바르게는 이미지는 색상·질감·형태 등의 자질로 색인한다.
ㄴ X: '작곡가'는 음악 자체에서 추출한 신호 자질이 아니라 음악을 설명하는 메타데이터이다. 올바르게는 음악은 멜로디·리듬·음높이 등 음악 내용의 특징으로 색인한다.
ㄷ O: 오디오의 높낮이, 크기, 발음은 음성·음향 신호의 내용에서 파악할 수 있는 자질이다.
ㄹ O: 비디오의 색분포, 질감, 돌출점은 프레임의 시각적 내용을 분석하여 얻는 자질이다.

선지별 해설

① X: ㄱ과 ㄴ은 모두 내용 자질이 아닌 '이미지 유형'과 '작곡가'를 포함한다. 올바른 조합은 ㄷ, ㄹ이다.
② X: ㄷ은 옳지만 ㄱ이 틀리고, 옳은 ㄹ이 빠져 있다. 올바른 조합은 ㄷ, ㄹ이다.
③ X: ㄹ은 옳지만 ㄱ이 틀리고, 옳은 ㄷ이 빠져 있다. 올바른 조합은 ㄷ, ㄹ이다.
④ X: ㄷ은 옳지만 ㄴ이 틀리고, 옳은 ㄹ이 빠져 있다. 올바른 조합은 ㄷ, ㄹ이다.
⑤ O: ㄷ의 오디오 자질과 ㄹ의 비디오 자질이 모두 멀티미디어 자체의 내용에서 추출되는 특징이므로 정확한 조합이다.

NoteBase

오답 포인트

ㄱ의 '대상·크기'와 ㄴ의 '멜로디·연주악기'처럼 그럴듯한 항목만 보고 묶음 전체를 옳다고 판단하기 쉽다. 그러나 ㄱ에는 '이미지 유형', ㄴ에는 '작곡가'라는 기술 정보가 섞여 있으므로 각 묶음의 모든 항목을 확인해야 한다.

다시 나오면 이렇게 판단

해당 정보가 멀티미디어 신호나 화면을 분석하여 얻어지는지 먼저 판단한다. 색상·질감·형태·음높이·음의 크기처럼 내용에서 추출되면 내용 자질이고, 작곡가·파일 유형처럼 자료를 설명하기 위해 별도로 붙인 정보이면 메타데이터이다. 복수 항목 묶음은 하나라도 내용 자질이 아니면 전체를 틀린 것으로 판단한다.

암기 포인트

내용기반 색인은 '자료 안에서 추출', 메타데이터는 '자료 밖에서 설명'이다. 이 문항에서는 오디오 ㄷ과 비디오 ㄹ이 옳다.

NoteBase

17. 다음 설명에 부합하는 용어는?

> 온라인 상의 텍스트, 이미지, 동영상, 음성 파일 등에 특정한 코드값을 붙여 관리하도록 고안된 코드로 저작물의 정보를 쉽게 식별하고 추적할 수 있게 한 표준정보 식별자이다. URL의 단점을 보완하는 기능에 추가적으로 문서 자동추적 기능, 워터마킹 기능 등 다양한 정보서비스를 추가할 수 있는 융통성 있는 번호체계이다. 미국출판협회(AAP) 등이 디지털 저작물에 대한 지적재산권 보호를 위해 처음 도입하였다.

① RDF(Resource Description Framework)
② DOI(Digital Object Identifier)
③ UCI(Universal Content Identifier)
④ Open URL(Uniform Resource Locator)
⑤ URI(Uniform Resource Identifier)

정답 2번

출제 포인트

이 문제는 디지털 저작물에 영구적인 식별자를 부여하는 DOI와 RDF, UCI, OpenURL, URI의 기능 차이를 묻는다.

핵심 개념

DOI(Digital Object Identifier)는 디지털 객체에 고유하고 지속적인 식별자를 부여하고, 연결 정보가 바뀌더라도 객체를 식별·추적할 수 있게 하는 체계이다. 설명문의 '특정 코드값', '식별과 추적', 'URL의 단점 보완', 'AAP가 디지털 저작물의 지식재산권 보호를 위해 도입'이 DOI의 결정적 단서이다.
RDF는 웹 자원의 의미와 관계를 주어-술어-목적어 구조로 표현하는 메타데이터 프레임워크이고, URI는 자원을 식별하는 문자열의 포괄적 체계이다. UCI는 국내 디지털 콘텐츠에 식별 코드를 부여하는 콘텐츠 식별체계이며, OpenURL은 이용자의 소속기관과 이용 가능 자원 같은 문맥 정보를 바탕으로 적절한 서비스로 연결하는 링크 표준이다.

NoteBase

선지별 해설

① X — RDF는 자원의 속성과 자원 사이의 관계를 표현하는 메타데이터 기술이다. 틀린 부분은 '저작물에 영구적인 코드값을 부여하여 식별·추적하는 표준 식별자'로 본 점이며, 올바르게는 그 개념이 DOI이다.
② O — DOI는 디지털 객체마다 지속적인 식별자를 부여하여 객체의 위치가 변경되어도 식별과 연결을 유지하도록 한다. AAP가 디지털 저작물의 지식재산권 보호를 위해 도입했다는 설명도 DOI와 일치한다.
③ X — UCI는 국내 디지털 콘텐츠에 고유 코드를 부여하여 유통과 관리를 지원하는 식별체계이다. 틀린 부분은 AAP가 처음 도입한 국제적 디지털 객체 식별체계로 본 점이며, 올바르게는 DOI가 이에 해당한다.
④ X — OpenURL은 이용자의 문맥 정보를 포함해 적절한 원문이나 도서관 서비스로 연결하는 개방형 링크 표준이다. 틀린 부분은 저작물 자체에 영구 코드값을 부여하는 식별체계로 본 점이며, 올바르게는 DOI가 그 역할을 한다.
⑤ X — URI는 인터넷상의 자원을 식별하는 문자열을 통칭하며 URL도 URI의 한 유형이다. 틀린 부분은 AAP가 저작권 보호를 위해 도입하고 부가 정보서비스를 제공하는 특정 디지털 저작물 식별체계로 본 점이며, 올바르게는 DOI이다.

오답 포인트

DOI와 URI는 모두 자원을 식별한다는 점 때문에 혼동하기 쉽다. 그러나 이 문항의 'AAP', '디지털 저작물', '영구적인 코드값', 'URL의 단점 보완'이라는 표현은 일반적인 식별자 범주인 URI가 아니라 DOI를 특정한다.

다시 나오면 이렇게 판단

영구 식별자와 위치 변경 후에도 유지되는 연결을 강조하면 DOI로 판단한다. 자원 간 의미 관계는 RDF, 국내 콘텐츠 식별 코드는 UCI, 문맥 기반 링크 연결은 OpenURL, 인터넷 자원 식별자의 포괄 개념은 URI로 구분한다.

암기 포인트

DOI는 디지털 객체의 영구 식별, URI는 자원 식별의 큰 범주, RDF는 관계 표현, OpenURL은 문맥 기반 연결이다.

NoteBase

18. 다음 <보기>의 설명이 지칭하는 프로토콜을 각각 옳게 나열한 것은?

<보 기>

ㄱ. 클라이언트와 호스트 데이터베이스 관리시스템 사이에 표준 인터페이스를 이용하여 서지정보를 검색하는 데 사용함

ㄴ. 네트워크 출판 프로토콜과 인터넷을 통해 데이터베이스를 검색하는 데 사용함

ㄷ. Z39.50에 기반한 NISO의 제안으로 19개의 대화명령어를 정의하였으며 고전적 불리언 모델에 기반함

	ㄱ	ㄴ	ㄷ
①	Z39.50	WAIS	CCL
②	Z39.50	OPAC	CCL
③	Z39.50	CCL	WAIS
④	OPAC	Z39.50	CCL
⑤	CCL	Z39.50	WAIS

* OPAC(Online Public Access Catalog)
* CCL(Common Command Language)
* WAIS(Wide Area Information Service)

정답 1번

출제 포인트

서지정보 검색과 관련된 Z39.50, WAIS, CCL의 정의와 특징을 구별하여 보기 ㄱ~ㄷ에 대응시키는 문제이다.

핵심 개념

Z39.50: 서로 다른 시스템의 클라이언트와 호스트 데이터베이스 관리시스템 사이에서 표준 인터페이스를 통해 서지정보를 검색하게 하는 정보검색 프로토콜이다. 키워드는 '클라이언트-서버 간 표준 인터페이스'이다.
WAIS: 네트워크 출판 프로토콜을 이용하여 인터넷상의 데이터베이스를 검색하는 분산 정보검색 방식이다. 키워드는 '인터넷 데이터베이스 검색'이다.
CCL: 19개의 대화명령어를 정의하고 고전적 불리언 모델에 기초한 공통 명령 언어이다. 키워드는 '대화명령어와 불리언 모델'이다.
OPAC: 이용자가 온라인으로 도서관 소장자료의 서지정보를 검색하는 공개 접근 목록이다. 이 보기의 세 프로토콜 설명 중 어느 것에도 해당하지 않는다.
따라서 ㄱ은 Z39.50, ㄴ은 WAIS, ㄷ은 CCL이다.

NoteBase

선지별 해설

① O — ㄱ=Z39.50, ㄴ=WAIS, ㄷ=CCL로 세 설명을 모두 옳게 연결한다.
② X — ㄴ을 OPAC으로 연결한 부분이 틀렸다. 인터넷을 통해 데이터베이스를 검색하는 네트워크 출판 프로토콜은 WAIS이므로, 올바르게는 'ㄱ Z39.50, ㄴ WAIS, ㄷ CCL'이다.
③ X — ㄴ과 ㄷ의 연결이 서로 바뀌었다. ㄴ은 WAIS이고, 19개의 대화명령어와 고전적 불리언 모델을 특징으로 하는 ㄷ은 CCL이므로, 올바르게는 'ㄱ Z39.50, ㄴ WAIS, ㄷ CCL'이다.
④ X — ㄱ을 OPAC, ㄴ을 Z39.50으로 연결한 부분이 틀렸다. 표준 인터페이스를 이용하는 ㄱ은 Z39.50이고 인터넷 데이터베이스 검색에 쓰이는 ㄴ은 WAIS이므로, 올바르게는 'ㄱ Z39.50, ㄴ WAIS, ㄷ CCL'이다.
⑤ X — 세 연결이 모두 틀렸다. ㄱ은 CCL이 아니라 Z39.50, ㄴ은 Z39.50이 아니라 WAIS, ㄷ은 WAIS가 아니라 CCL이므로, 올바르게는 'ㄱ Z39.50, ㄴ WAIS, ㄷ CCL'이다.

오답 포인트

Z39.50과 WAIS가 모두 데이터베이스 검색에 관련된다는 점 때문에 ㄱ과 ㄴ을 바꾸기 쉽고, OPAC도 온라인 검색 도구라는 이유로 프로토콜과 혼동하기 쉽다. '표준 인터페이스'는 Z39.50, '인터넷 데이터베이스'는 WAIS, '19개 명령어·불리언'은 CCL로 구분한다.

다시 나오면 이렇게 판단

설명의 결정적 단서를 먼저 찾는다. 클라이언트와 호스트 사이의 표준 인터페이스이면 Z39.50, 네트워크 출판과 인터넷 데이터베이스 검색이면 WAIS, 대화명령어와 불리언 검색 언어이면 CCL로 판단한다. OPAC은 프로토콜 명칭이 아니라 이용자용 온라인 공개 목록이라는 점을 함께 구별한다.

암기 포인트

표준 인터페이스는 Z39.50, 인터넷 분산검색은 WAIS, 19개 명령어와 불리언은 CCL, 온라인 공개목록은 OPAC이다.

NoteBase

19. 다음 <보기>의 검색시스템이 등장한 시대 순서로 옳게 나열한 것은?

<보 기>

ㄱ. 링킹 시스템(linking system)
ㄴ. 본문 검색시스템(full-text retrieval system)
ㄷ. 오프라인 배치시스템(offline batch system)
ㄹ. 유니텀 시스템(uniterm system)
ㅁ. 주변절단카드 시스템(edge-notched card system)

① ㄷ → ㄹ → ㅁ → ㄴ → ㄱ
② ㄷ → ㅁ → ㄹ → ㄱ → ㄴ
③ ㄹ → ㅁ → ㄷ → ㄴ → ㄱ
④ ㅁ → ㄴ → ㄷ → ㄱ → ㄹ
⑤ ㅁ → ㄹ → ㄷ → ㄴ → ㄱ

정답 5번

출제 포인트

검색시스템이 수작업 카드 방식에서 전산화·온라인화되고, 문헌 간 연결을 활용하는 방식으로 발전한 순서를 묻는 문제이다.

올바른 순서

ㅁ. 주변절단카드 시스템 → ㄹ. 유니텀 시스템 → ㄷ. 오프라인 배치시스템 → ㄴ. 본문 검색시스템 → ㄱ. 링킹 시스템

1. 주변절단카드 시스템 — 카드 가장자리의 절단·천공 상태를 이용하여 자료를 수작업으로 선별하는 초기 검색 방식이다.
2. 유니텀 시스템 — 문헌의 내용을 단일 개념어인 유니텀으로 나타내고 개념어들을 조합하여 검색하는 좌표색인 방식이다.
3. 오프라인 배치시스템 — 검색 요구를 모아 두었다가 컴퓨터가 일괄 처리하는 초기 전산 검색 방식이다.
4. 본문 검색시스템 — 문헌의 색인어뿐 아니라 전자화된 본문 자체에 포함된 단어를 검색하는 방식이다.
5. 링킹 시스템 — 문헌과 문헌을 연결한 링크를 따라 관련 정보를 탐색하는 방식이다.

왜 이 순서인가: 검색 매체와 처리 방식이 물리적 카드의 수작업 선별에서 개념어 조합, 컴퓨터 일괄 처리, 전자 본문 검색, 연결된 문헌 탐색의 순서로 발전했기 때문이다.

NoteBase

선지별 해설

① X — ㄷ의 오프라인 배치시스템을 ㄹ의 유니텀 시스템과 ㅁ의 주변절단카드 시스템보다 앞에 둔 점이 틀렸다. 올바르게는 ㅁ → ㄹ → ㄷ → ㄴ → ㄱ이다.

② X — 전산식인 ㄷ의 오프라인 배치시스템을 초기 수작업 카드 방식인 ㅁ보다 앞에 둔 점이 틀리고, ㄱ의 링킹 시스템과 ㄴ의 본문 검색시스템의 선후도 뒤바뀌었다. 올바르게는 ㅁ → ㄹ → ㄷ → ㄴ → ㄱ이다.

③ X — ㄹ의 유니텀 시스템을 더 이른 ㅁ의 주변절단카드 시스템보다 앞에 둔 점이 틀렸다. 올바르게는 ㅁ → ㄹ → ㄷ → ㄴ → ㄱ이다.

④ X — ㅁ을 처음에 둔 것은 맞지만, 곧바로 ㄴ의 본문 검색시스템을 배치하고 가장 오래된 축에 속하는 ㄹ의 유니텀 시스템을 마지막에 둔 점이 틀렸다. 올바르게는 ㅁ → ㄹ → ㄷ → ㄴ → ㄱ이다.

⑤ O — 초기의 주변절단카드 시스템에서 유니텀 시스템, 오프라인 배치시스템, 본문 검색시스템을 거쳐 링킹 시스템으로 발전한 순서가 정확하다.

오답 포인트

유니텀 시스템을 컴퓨터 기반의 최신 검색 방식으로 오해하거나, 본문 검색과 링킹의 순서를 바꾸기 쉽다. 유니텀은 단일 개념어를 조합하는 초기 좌표색인 방식이고, 본문 검색 다음에 문헌 간 연결을 활용하는 링킹 시스템이 등장한다.

다시 나오면 이렇게 판단

검색시스템의 시대 순서는 처리 매체와 검색 범위를 기준으로 판단한다. 물리적 카드 → 개념어 조합 → 컴퓨터 일괄 처리 → 전자 본문 → 문헌 간 링크의 흐름을 적용한다.

암기 포인트

주변절단카드-유니텀-배치-본문-링크, 즉 '카드에서 링크까지'의 흐름으로 기억한다.

NoteBase

20. 웹 검색 결과의 품질을 평가하는 방식 또는 척도에 대한 설명으로 옳은 것은?

① 품질 평가에 클릭데이터를 사용하는 장점은 정확성이며 단점은 클릭의 확장 제한성이다.

② 클릭률은 랭킹, 사용자 인터페이스와 독립적이다.

③ 웹에서 재현율을 측정하는 것은 거의 불가능하기에 정확률과 재현율 숫자를 직접 사용할 수 없다.

④ 검색 스니펫(snippets)은 각 검색 결과에 대한 이용자 평가를 시스템에 알려 주는 피드백 기능을 한다.

⑤ 검색 스니펫(snippets)은 페이지 랭킹에 영향을 미치기 때문에 검색 결과의 품질 평가에 사용될 수 없다.

정답 3번

출제 포인트

이 문제는 웹 검색 결과의 품질을 평가할 때 클릭데이터, 재현율·정확률, 검색 스니펫이 지니는 의미와 한계를 구분하는 문제이다.

핵심 개념

클릭데이터는 이용자가 어떤 검색 결과를 클릭했는지 기록한 암묵적 피드백이다. 많은 이용자의 행동을 수집할 수 있다는 장점이 있지만, 클릭은 결과의 순위와 화면 구성 등에 영향을 받으므로 관련성을 그대로 나타내는 정확한 평가값은 아니다.

정확률은 검색된 문서 중 관련 문서의 비율이고, 재현율은 전체 관련 문서 중 검색된 문서의 비율이다. 실제 웹에서는 문서와 검색 결과가 계속 바뀌고 전체 관련 문서 집합을 확정하기도 어려워, 동일한 조건을 재현하여 이 수치를 직접 측정하기가 매우 어렵다.

검색 스니펫은 검색 결과 페이지에 제시되는 문서의 제목·요약 등의 미리보기 정보이다. 이용자의 관련성 판단과 클릭에 영향을 주지만, 그 자체가 이용자 평가를 시스템에 전달하는 피드백 기능은 아니다.

NoteBase

선지별 해설

① X — 틀린 표현은 클릭데이터의 장점을 '정확성'이라고 한 부분과 단점을 '클릭의 확장 제한성'이라고 한 부분이다. 클릭데이터의 장점은 대규모 이용자 행동을 비교적 쉽게 축적할 수 있다는 점이고, 단점은 순위·화면 구성 등의 편향과 잡음이 섞여 관련성을 정확히 나타내지 못한다는 점이다.
② X — 틀린 표현은 '클릭은 랭킹, 사용자 인터페이스와 독립적이다'이다. 실제 클릭은 상단에 노출된 결과가 더 자주 선택되는 순위 편향과 제목·스니펫 등 인터페이스의 영향을 받으므로 이들과 독립적이지 않다.
③ O — 실제 웹은 문서와 검색 결과가 계속 변하고 전체 관련 문서 집합을 완전하게 정하기도 어렵다. 따라서 동일 조건에서 재현율을 측정하는 것은 거의 불가능하며, 정확률과 재현율 수치를 실제 웹 검색 품질 평가에 직접 적용하기 어렵다는 설명은 옳다.
④ X — 틀린 표현은 스니펫이 이용자 평가를 시스템에 알려 주는 '피드백 기능'을 한다는 부분이다. 스니펫은 문서 내용을 미리 보여 주어 이용자의 관련성 판단과 클릭을 돕는 결과 제시 요소이며, 이용자의 평가는 클릭 등의 별도 행동을 통해 간접적으로 관찰된다.
⑤ X — 틀린 표현은 스니펫이 '페이지 랭킹에 영향을 미치기 때문에' 품질 평가에 사용될 수 없다는 부분이다. 스니펫은 검색 결과가 제시될 때 이용자의 판단과 클릭에 영향을 줄 수 있으므로 클릭데이터 해석에서 고려해야 할 요소이며, 그 영향 때문에 품질 평가가 불가능해지는 것은 아니다.

오답 포인트

③의 '거의 불가능'이라는 강한 표현 때문에 틀렸다고 판단하기 쉽다. 그러나 실제 웹 전체를 대상으로 하면 관련 문서의 총집합을 확정하기 어렵고 검색 환경도 계속 변하므로 재현율을 직접 측정하기 어렵다. 또한 스니펫이 클릭을 유도한다는 사실과 스니펫 자체가 피드백 기능이라는 주장은 구분해야 한다.

다시 나오면 이렇게 판단

평가 방법을 볼 때 먼저 무엇을 측정하는지 확인한다. 클릭은 이용자 행동 기록이지만 순위와 인터페이스의 영향을 받고, 정확률·재현율은 관련 문서 집합이 필요하며, 스니펫은 결과를 보여 주는 인터페이스 요소이다. 이 세 역할을 구분하면 선지의 속성 바꾸기를 판별할 수 있다.

암기 포인트

클릭은 대규모이지만 편향되고, 재현율은 전체 관련 문서 집합이 필요하며, 스니펫은 피드백이 아니라 결과 미리보기이다.

NoteBase