

정보학개론

1. 다음 설명에 해당하는 학술지 영향력을 평가하는 지표는?

- 한 학술지에 5년 동안 출판된 총 논문 수와 피인용횟수를 기반으로 측정한다.
- 한 학술지가 다른 학술지에 비해 갖는 상대적 영향력을 평가하는 데 사용한다.
- 페이지랭크 알고리즘(PageRank algorithm)과 유사한 방식으로 계산한다.

- ① SNIP(Source Normalized Impact per Paper)
- ② Eigenfactor
- ③ JIF(Journal Impact Factor)
- ④ CiteScore
- ⑤ Cited Half-Life

2. 다음 설명이 지칭하는 에뮬레이션(emulation) 방식을 각각 옳게 나열한 것은?

- ㄱ. 원본에 접근하는 데 요구되는 모든 기술(S/W, H/W, O/S 등)을 보존하는 방식이다.
- ㄴ. 디지털 원본에 적용된 기술적인 조건들에 변경이 있어도 인코딩되어 있는 콘텐츠를 재생할 수 있는 환경을 프로그램으로 만들어 내어 디지털 문헌의 접근성을 보장하는 방식이다.
- ㄷ. 디지털 파일의 내용 해석을 UVC(Universal Virtual Computer)의 기계언어로 처리할 수 있도록 프로그램을 작성하여 해결하려는 시도이다.

- | | ㄱ | ㄴ | ㄷ |
|---|-----------|-----------|-----------|
| ① | 기술보존 | 기술 에뮬레이션 | 가상기계 프로그램 |
| ② | 기술보존 | 가상기계 프로그램 | 기술 에뮬레이션 |
| ③ | 가상기계 프로그램 | 기술 에뮬레이션 | 기술보존 |
| ④ | 가상기계 프로그램 | 기술보존 | 기술 에뮬레이션 |
| ⑤ | 기술 에뮬레이션 | 기술보존 | 가상기계 프로그램 |

3. 자연어 색인의 장점으로 옳은 것만을 <보기>에서 모두 고르면?

- < 보 기 > —
- ㄱ. 다양한 접근점의 제공이 가능하다.
 - ㄴ. 새로운 개념의 표현이 용이하다.
 - ㄷ. 용어의 계층관계 표시가 수월하다.
 - ㄹ. 융통성과 표현력이 좋다.
 - ㅁ. 동의어, 동형이의어의 통제가 쉽다.

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

4. 다음은 연구자 A의 논문을 피인용횟수가 많은 순으로 정렬한 표이다. 연구자 A의 h지수는?

논문명	피인용순위	피인용횟수
가	1	45
나	2	30
다	3	24
라	4	15
마	5	14
바	6	11
사	7	9
아	8	7
자	9	6
차	10	5

- ① 5
- ② 6
- ③ 7
- ④ 8
- ⑤ 9

5. HTML과 XML에 대한 설명으로 옳은 것만을 <보기>에서 모두 고르면?

< 보 기 >

- ㄱ. HTML은 W3C(World Wide Web Consortium) 표준이다.
 ㄴ. HTML은 이종의 컴퓨터들이 처리할 수 있는 웹 문서를 작성하는 언어이다.
 ㄷ. HTML과 XML 모두 태그(tag)를 확장하여 사용 가능하다.
 ㄹ. XML은 국제표준이며 상이한 응용 프로그램 간의 호환을 가능하게 한다.
 ㅁ. HTML은 데이터 구조를 기술하는 기능에 한계가 없다.

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

6. 불리언 모델에서 문헌의 색인어뿐만 아니라 질의어에도 가중치를 부여하는 방법으로 옳은 것은?

- ① 퍼지집합 모델
 ② MMM 모델
 ③ P-Norm 모델
 ④ 필드 가중치 기법
 ⑤ 논리합 정규형 기법

7. 다음 CCL(Creative Commons License)의 이용허락조건에 대한 설명으로 옳지 않은 것은?



- ① ㄱ은 '저작자 표시'로 저작물을 복사하거나 다른 곳에 게시할 때도 반드시 저작자와 출처를 표시해야 함을 의미한다.
 ② ㄴ은 '비영리'를 표시하며 저작물을 영리목적으로 이용하려면 별도의 계약이 필요함을 의미한다.
 ③ ㄷ은 '2차 저작물 표시'로 원 저작물을 변경한 2차 저작물일 경우에 기입해야 함을 의미한다.
 ④ ㄹ은 '동일조건변경허락'으로 2차 저작물 창작을 허용하되, 원 저작물과 동일한 라이선스를 적용해야 함을 의미한다.
 ⑤ 이용허락조건 4종류를 조합하여 총 6종류의 CCL이 존재한다.

8. BIBFRAME에 대한 설명으로 옳지 않은 것은?

- ① 핵심 상위 클래스인 인스턴스는 저자, 언어, 주제를 포함한다.
 ② 서지정보를 시멘틱웹에서 사용할 수 있도록 링크드데이터 형식으로 구축하고 RDF/XML로 인코딩한다.
 ③ 부가적 클래스인 이벤트는 저작의 내용과 관련된 사건이다.
 ④ 개별자료 클래스는 인스턴스의 실질적 사본의 물리적 혹은 가상의 소장위치를 포함한다.
 ⑤ 부가적 클래스인 에이전트는 개인, 단체, 가족, 정부 등을 포함한다.

9. 다음 문헌분리값 공식에 기반한 설명으로 옳지 않은 것은?

$$\text{문헌분리값}_k = Q_k - Q$$

문헌분리값_k : 단어 k의 문헌분리값 Q_k : 단어 k를 제거한 후의 평균 유사도 Q : 단어 k를 색인어로 부여하였을 때 평균 유사도

- ① 단어의 문헌분리값은 특정한 단어가 한 문헌집단에서 상호 관련 없는 문헌을 분리하는 능력을 측정한 것이다.
 ② 문헌분리값이 큰 단어일수록 좋은 색인어가 된다.
 ③ 단어 k가 주제가 다른 문헌을 가능한 한 분리시키는 좋은 색인어라면 $Q_k - Q$ 는 양수가 된다.
 ④ 단어 k가 많은 문헌에 출현하는 경우 $Q_k - Q$ 는 음수가 될 수 있다.
 ⑤ 단어 k의 가중치는 문헌빈도와 문헌분리값을 곱하여 계산한다.

10. 텍스트 마이닝에 대한 설명으로 옳지 않은 것은?

- ① 자연어 문서에 수록된 비정형 텍스트 데이터를 분석하여 어떤 규칙이나 패턴 등 유용한 정보를 찾는 것이다.
 ② 대소문자 통일, 문장부호·특수문자 제거, 불용어 삭제 등의 데이터 전처리 과정이 필요하다.
 ③ 감성분석은 단어와 문맥을 분석하여 텍스트에 들어 있는 사람들의 태도, 의견, 성향과 같은 주관적인 데이터를 분석하는 기술이다.
 ④ 단어의 출현빈도를 분석하는 단어 구름(word cloud)은 핵심 단어를 추출하여 문서의 주제를 객관적이면서 명확하게 식별하는 것이다.
 ⑤ 주요 응용분야로는 정보 추출, 문서 요약, 문서 분류 등이 있으며, 부가적으로 분석결과에 대한 정보 시각화를 수행하기도 한다.

11. 다음에서 설명하는 메타데이터 표준에 정의된 요소명이 아닌 것은?

- MARC, ONIX, TEI 등을 절충하여 상호운용성과 정밀성을 만족시키도록 설계되었다.
- 디지털 도서관 영역의 범용 서지정보 메타데이터로 MARC와 상호호환이 가능하다.
- XML 스키마 기반의 인코딩 구조를 채택하며 언어형식의 태그를 사용한다.

- ① typeOfResource
- ② accessCondition
- ③ identifier
- ④ title
- ⑤ originInfo

12. OAK(Open Access Korea) 리포지터리에 대한 설명으로 옳은 것만을 <보기>에서 고르면 모두 몇 개인가?

- < 보 기 > —
- ㄱ. 한국형 오픈엑세스 리포지터리로 Google 등 해외 검색포털 등과의 연계가 불가하다.
 - ㄴ. 국제적으로 많이 사용하는 DSpace를 기반으로 개발되었다.
 - ㄷ. 공공기관, 대학, 연구소 등의 국내 기관에서 생산한 디지털 지식정보를 보존 및 관리한다.
 - ㄹ. 다양한 포맷(PDF, HWP, DOC 등)의 텍스트 자료만을 아카이빙하고 있다.
 - ㅁ. 현재 국립중앙도서관이 국가지식 정보 구축 및 확산 사업으로 추진하고 있다.

- ① 1개
- ② 2개
- ③ 3개
- ④ 4개
- ⑤ 5개

13. 정보검색 모델에 대한 설명으로 옳지 않은 것은?

- ① LSI(Latent Semantic Indexing) 모델은 대수적(algebraic) 모델에 속하는 개념색인 기법이다.
- ② 확장 불리언 모델은 용어의 관계를 설정할 수 있지만 부분일치 검색은 허용하지 않는다.
- ③ 벡터공간 모델은 대수적(algebraic) 모델로 부분일치 검색과 순위화가 가능하다.
- ④ 확률 모델은 베이즈 정리(Bayes' theorem)를 변형하여 적합성 확률을 구할 수 있다.
- ⑤ 퍼지집합 모델은 한 문헌이 특정한 색인어로 표현되는 개념 범주에 속하는 정도를 나타내는 부분 멤버십을 이용한다.

14. 다음 <조건>에 따라 계산한 정보검색 평가척도의 값으로 옳은 것은?

- < 조 건 > —
- ㄱ. 검색 대상 전체 문헌 수: 100개
 - ㄴ. 검색된 문헌 수: 20개
 - ㄷ. 검색된 문헌 중 검색자 요구에 적합한 문헌 수: 8개
 - ㄹ. 검색되지 않은 문헌 중 검색자 요구에 적합한 문헌 수: 60개

- ① 정확률은 20.0%이다.
- ② 재현율은 7.5%이다.
- ③ 잡음률은 80.0%이다.
- ④ 부적합률은 37.5%이다.
- ⑤ 누락률은 92.5%이다.

15. n-그램(n-gram) 색인에 대한 설명으로 옳지 않은 것은?

- ① 형태소 분석이나 스테밍과 같은 언어학적 처리가 필요 없다.
- ② 띄어쓰기를 사용하지 않는 아시아권 문자를 색인하는 데 유용하다.
- ③질의어를 n-그램으로 변환하여 용어절단 검색을 할 수 있다.
- ④질의어와 색인어에 철자 오류가 있더라도 유사도가 높은 색인어를 검색할 수 있다.
- ⑤ 특정 언어의 문법에 독립적이기 때문에 색인의 크기가 줄어든다.

16. with 연산자를 사용하는 검색시스템에서 색인어 A와 B가 다음과 같이 문헌에 출현하는 경우, 검색식 ‘A with B’로 검색되는 문헌은?

색인어	문헌번호	출현위치
A	1	7
	2	5
	3	2
	4	4
B	1	6
	2	6
	4	10
	5	19

- ① 1번 문헌
- ② 2번 문헌
- ③ 3번 문헌
- ④ 4번 문헌
- ⑤ 5번 문헌

17. 데이터 관리와 서비스를 위해 데이터 사서가 관심 가져야 할 요소로 옳지 않은 것은?

- ① 기관형 데이터 리포지터리
- ② 데이터 아카이빙 정책개발
- ③ 데이터와 문헌 연계 서비스
- ④ DMP(Data Management Plan) 작성지원 및 컨설팅
- ⑤ 데이터 이산화(discretization) 컨설팅

18. 베이츠(M. Bates)는 정보학의 핵심 연구영역에서 ‘세 가지 큰 질문(Three Big Questions)’을 던지고 있다고 설명하였다. 다음 중 ‘세 가지 큰 질문’의 유형과 설명이 옳게 연결된 것은?

- ① 물리적 질문 - 기록된 정보세계의 특징과 법칙은 무엇인가?
- ② 설계 질문 - 사람들은 어떻게 정보를 찾고, 이해하고, 이용하는가?
- ③ 사회적 질문 - 정보의 사회적 효용성과 법적 가치는 무엇인가?
- ④ 물리적 질문 - 기록된 정보에 대한 가장 빠르고 효과적인 접근은 어떻게 이루어질 수 있는가?
- ⑤ 사회적 질문 - 정보사회를 어떻게 기술하고 설명하는가?

19. 클러스터링에 대한 설명으로 옳지 않은 것은?

- ① 비계층적 클러스터링 알고리즘은 미리 생성할 클러스터의 기준치를 결정하므로 경험적 기법이라고도 한다.
- ② 단일연결 기법은 계층적 클러스터링 알고리즘으로 통합 대상이 되는 두 클러스터에서 가장 유사한 두 개 구성원의 유사도를 기반으로 클러스터를 생성하는 기법이다.
- ③ 계층적 클러스터링 알고리즘은 처리 시간 측면에서 비계층적 기법에 비하여 효율성이 떨어진다.
- ④ 완전연결 기법은 계층적 클러스터링 알고리즘으로 통합 대상이 되는 두 클러스터에서 가장 유사하지 않은 두 개 구성원의 유사도를 기반으로 클러스터를 생성하는 기법이다.
- ⑤ K-means 클러스터링 기법은 계층적 클러스터링 알고리즘으로 임의로 형성된 K개의 초기 클러스터 집합에서 시작된다.

20. 정보검색 시스템의 성능 평가척도들 중 특정 질의에 대한 전체 적합 문헌 수를 알지 못하더라도 사용 가능한 척도만을 <보기>에서 모두 고르면?

< 보 기 >	
ㄱ. 정확률	ㄴ. 재현율
ㄷ. 잡음률	ㄹ. 누락률
ㅁ. E 척도	ㅂ. F 척도
ㅅ. R-정확률(R-precision)	ㅇ. 순위정확률(precision at k)

- ① ㄱ, ㄷ, ㅁ
- ② ㄱ, ㄷ, ㅅ
- ③ ㄱ, ㄷ, ㅇ
- ④ ㄴ, ㄹ, ㅂ
- ⑤ ㄴ, ㄹ, ㅅ, ㅇ