

# 2020年度 検索技術者検定

## 2級 試験問題(前半)

### 注意事項

1. 着席したら、受験票を机の上に置いて下さい。
2. 解答用紙の所定の欄に受験番号・氏名を必ずご記入下さい。
3. 解答時間は、13：40～15：10の90分間です。
4. 中途退席できる時間は、14：25～14：55の30分間です。  
一旦退席すると、前半の試験終了まで再入場はできません。
5. 問題は12問、全13ページ、解答用紙は2ページです。確認の上、落丁・乱丁・印刷不鮮明のもの等がありましたら、手をあげて試験官にお知らせ下さい。
6. 解答は、問題文の指示にしたがい、解答用紙にご記入下さい。  
(解答用紙裏面への記入は無効です。)
7. 問題の内容に関する質問は一切できません。
8. 試験問題は持ち帰って結構です。

問1 以下は情報検索技術に関する記述である。空欄 (A) ~ (G) にあてはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

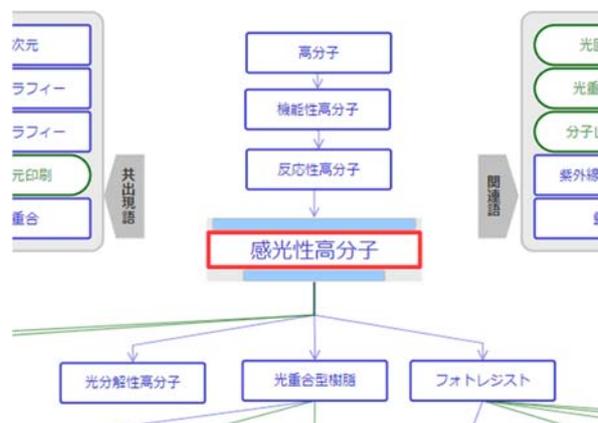
- (1) 次の文書 ID=1 と文書 ID=2 を検索可能なインデックスファイルを構築する場合に、助詞を (A) として (B) 解析を行い、索引語を抽出した。文書 ID と文書における索引語の出現位置を格納してできた次のファイルを (C) という。

文書 ID=1 「人文 科学 分野 の 情報 資源 と 検索」  
 文書 ID=2 「ライフ サイエンス 分野 の 情報 資源 と 検索」

単語	(文書 ID, 出現位置)
人文	(1, 1)
科学	(1, 2)
分野	(1, 3)、(2, 3)
情報	(1, 4)、(2, 4)
資源	(1, 5)、(2, 5)
検索	(1, 6)、(2, 6)
ライフ	(2, 1)
サイエンス	(2, 2)

- (2) (D) は全ての検索システムが同一の機能を有しているとは限らないため注意が必要であるが、例えば入力した語順に指定して入力語の間に2語以内の語を含む隣接は (2W) で表され、語順は問わないが2語以内の語を含む隣接は (2N) で表される。たとえば、「システムを正常に働かせる」という内容が含まれる文章を検索するとき、表現「get the system to work properly」が考えられる。このような表現を検索するには、(D) の検索条件を (E) とすることが考えられる。

- (3) 次の図は、JSTソーラス map を用いて「感光性高分子」の検索した結果の一部を示すものである。光分解性高分子、光重合型樹脂、およびフォトレジストは、「感光性高分子」の (F) である。また、「感光性高分子」は (G) であるから、「感光性高分子」を検索に用いると「感光性高分子」の類語・同義語を含めて検索できる。



<解答群>

- |                          |                       |                          |
|--------------------------|-----------------------|--------------------------|
| 1. シーケンシャルファイル           | 2. 統制語                | 3. Narrower term         |
| 4. 形態素                   | 5. 近接演算               | 6. ストップワード               |
| 7. 類義語                   | 8. 論理演算               | 9. ユニターム索引               |
| 10. 上位語                  | 11. 全文検索              | 12. TF・IDF               |
| 13. スコープノート              | 14. 下位語               | 15. 関連語                  |
| 16. 構文                   | 17. 用語抽出              | 18. 転置ファイル               |
| 19. properly (3N) system | 20. get (3W) properly | 21. properly (3W) system |

問2 以下は検索結果の評価に関する記述である。空欄(A)～(F)にあてはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

検索結果の評価方法には、顧客満足度を調査する方法と、(A)と(B)を算定する方法の2種類がある。(A)は、検索対象の集合中に存在する検索テーマに合致する適合情報のうち、実際に検索された適合情報の割合を表す。これはどの程度、(C)があったのかを示す指標になる。(B)は、実際に得た検索結果のうち、適合情報がどれだけ検索されたのか、その割合を示す。これは(D)を示す指標になる。下図では実際に検索された情報全体は $y+z$ であり、検索要求に合致する適合情報全体は $x+y$ である。下図において(A)は(E)と算出され、(B)は(F)と算出される。

	検索ノイズ z
検索漏れ x	検索された適合情報 y

<解答群>

- |                          |                          |                          |                         |
|--------------------------|--------------------------|--------------------------|-------------------------|
| 1. 統制語                   | 2. 精度                    | 3. 同義語                   | 4. 検索ノイズ                |
| 5. 再現率                   | 6. 検索漏れ                  | 7. $x/(x+y) \times 100$  | 8. $y/(y+z) \times 100$ |
| 9. $z/(x+z) \times 100$  | 10. $x/(x+z) \times 100$ | 11. $y/(x+y) \times 100$ |                         |
| 12. $z/(y+z) \times 100$ |                          |                          |                         |

問3 以下は、データベースの種類・流通に関する記述である。空欄（A）～（F）にあてはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

- (1) (A)とは、さまざまな電子化された情報資源を図書館のOPACと結びつけて利用者が検索に苦勞しなくても情報が得られる仕組みを提供するものである。
- (2) (B)は文献データベースともいい、文献の書誌情報や所在情報などを収録し、(C)へのアクセス情報を提供するデータベースである。
- (3) データベースから必要な情報を選択的に取り出せる（検索できる）よう、データベースの構築・管理を行う(D)を使って作られた情報検索サービスを提供するためのシステムを情報検索システムという。商用の情報検索システムの例には、JDreamⅢ、G-Search、日経テレコン、STNやDialogなどがある。
- (4) 自己のコンピュータを運用し、プロデューサからデータベースを委託され、またはデータベースプロデューサのデータベースに接続して、データベースをユーザに提供する機関を(E)あるいはベンダーという。
- (5) (F)は、一部大手の学術出版社による学術情報の寡占化と、学術情報の高騰に対抗して、査読付き学術誌に掲載された論文などを、学会や大学が自ら誰でも利用可能な形でウェブ上に置いたことに始まる。

<解答群>

- |                        |                 |             |
|------------------------|-----------------|-------------|
| 1. データベース管理システム (DBMS) | 2. レファレンスデータベース | 3. リンク・リゾルバ |
| 4. インフォメーションブローカー      | 5. オープンデータベース   | 6. オープンアクセス |
| 7. ディスカバリーサービス         | 8. ビッグディール契約    | 9. GUI      |
| 10. ディストリビュータ          | 11. インハウスデータベース | 12. プロバイダ   |
| 13. 零次情報               | 14. 一次情報        | 15. 二次情報    |

問4 以下は、書誌ユーティリティに関する記述である。(A)～(H)にあてはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

書誌ユーティリティの一つに、オンラインで多数の機関が1つのシステムに書誌データを登録する、(A)方式の目録作成のための組織がある。

我が国における代表的な書誌ユーティリティのシステムは、(B)が運用する、(C)である。

(C)は、対外的なサービスではあるが、目録担当者の業務用データベースとして運用を開始した。(C)は、2020年8月3日に、(D)がまとめた方針に従い、いわゆるCAT2020の運用を開始した。

エンドユーザが(C)のデータを検索できるシステムとしては、(B)の前身である学術情報センターが有償で提供した(E)が始まりであり、その後、汎用ウェブブラウザで検索可能な、無償で提供された(F)を経て、現在は(G)で検索が可能である。

また、(C)を用いた図書館間相互貸借のシステムとしては、(H)が現在も運用中である。

<解答群>

1. 国立国会図書館
2. 国立情報学研究所
3. 科学技術振興機構
4. 医学中央雑誌刊行会
5. これからの学術情報システム構築検討委員会
6. JOIS
7. NACSIS-CAT
8. NACSIS-IR
9. NACSIS-ILL
10. NACSIS-ELS
11. Webcat
12. GeNii
13. CiNii Articles
14. CiNii Books
15. CiNii Dissertations
16. CiNii Research
17. 共同分担
18. 一括集中

問5 以下は、検索システムやデータベースに関する説明である。(A)～(J)にあてはまる最も適切なシステム名またはデータベース名を解答群から選択し、その番号を解答欄に記入しなさい。

(A) EPO が発行した公報の全文と、EPO が収集した世界約 100 の国または特許機関が発行した公報の書誌事項と英文抄録を収録する。

(B) 金融庁が提供しているデータベースであり、上場企業および未上場有価証券報告書提出企業が開示した書類を検索・閲覧することができる。

(C) 米国電気電子学会によって制作された科学技術系の電子ジャーナルや電子ブックへのアクセスを提供するプラットフォーム。電気電子分野では世界有数のデータベースである。

(D) 科学技術振興機構が作成する科学技術や医学・薬学分野の国内外の書誌データベース。

(E) 目録所在情報サービスをもとにしたデータベースで、日本全国書誌 JAPAN/MARC (国立国会図書館) などから情報の提供を受けている。「連想検索」は、自然文の入力により関連した資料を検索することができる。

(F) 公的統計のポータルサイトで、各府省などが登録した統計データなどを集約して提供している。統計表ファイルを検索し、閲覧・ダウンロードすることができる。

(G) 日本に出願され、公報が発行された特許、実用新案、意匠、商標の情報を中心に収録している。非特許文献 (J-GLOBAL) の情報も検索が可能である。

(H) 総務省行政管理局が運営する総合的な行政情報ポータルサイト。法令検索や行政文書ファイル管理簿の検索などのサービスを提供している。

(I) Google が提供する図書検索サービス。書誌情報と本文の一部が表示され、著作権の状況によっては全文を閲覧することができる図書もある。

(J) 医学文献データベースとして国内最大であり、医学の他に薬学、歯学、看護学、獣医学に関する論文情報を収録している。「医学用語シソーラス」に体系的に収載された統制語が索引として付与されている。

<解答群>

- |                  |                                 |                  |                   |
|------------------|---------------------------------|------------------|-------------------|
| 1. NDL ONLINE    | 2. 外国特許情報サービス                   | 3. J-PlatPat     | 4. Google Scholar |
| 5. ELNET         | 6. JMEDPlus                     | 7. MathSciNet    | 8. MEDLINE        |
| 9. EDINET        | 10. Espacenet                   | 11. e-Stat       | 12. BIOSIS        |
| 13. PatentSQUARE | 14. IEEE Xplore Digital Library | 15. Google Books | 16. e-Gov         |
| 17. JDreamIII    | 18. 医中誌 Web                     | 19. 総務省統計局       | 20. FOPISER       |
| 21. STN          | 22. PATENTSCOPE                 | 23. EDGAR        | 24. Webcat Plus   |
| 25. JSTPlus      |                                 |                  |                   |

問6 以下は、ウェブアーカイビングに関する説明である。(A)～(H)にあてはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

ウェブページは、印刷物と異なり、その場の状況や伝えたいことを迅速に公開することを目的としていることが多いため、古い情報が残されず、常に新しい情報しか掲載されていないことが多い。

そこで、ウェブページをアーカイブし、なるべく公開した時と同じような形式・構造で保存するものがウェブアーカイビングである。

日本では、国立国会図書館が、インターネット資料収集保存事業(A)でアーカイブを行っている。(A)は(B)機関が作成するものが収集対象であり、(C)機関が作成するものは網羅的な収集としていない。

収集は、いくつかの技術を組み合わせて行っているが、ウェブアーカイブの国際コンソーシアム(D)が主導的な役割を果たしている。収集ロボットである(E)、全文検索エンジン、保存用ファイルフォーマット(F)、閲覧アプリケーション(G)などで構成される。

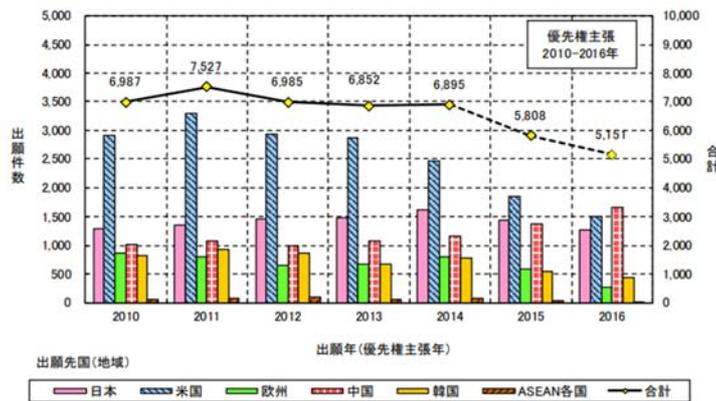
国外では、1996年にBrewster Kahle氏によって設立された、インターネットアーカイブが有名である。この非営利法人が運用する(G) Machineは、URLを指定した検索およびキーワード検索が可能であり、また日付指定で過去のウェブページを閲覧することが可能である。(G) Machineは無償であるが、有償のサービスである(H)も運用中である。

<解答群>

- |                |              |             |                |             |
|----------------|--------------|-------------|----------------|-------------|
| 1. NTFS        | 2. WARC      | 3. e デポ     | 4. Safari      | 5. 公的       |
| 6. クローラ        | 7. Wikipedia | 8. WIPO     | 9. WARP        | 10. Wayback |
| 11. IIPC       | 12. ローター     | 13. インデキシング | 14. Solr       | 15. arXiv   |
| 16. Archive-It | 17. 民間       | 18. chrome  | 19. デジタルコレクション |             |
| 20. 文書館        |              |             |                |             |

問7 以下は情報の利活用に関する記述である。(A)～(I)に当てはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

- (A) 大学や研究機関が所属する研究者の論文や研究成果を自ら蓄積するシステム。論文や研究成果は無料で利用できるオープンアクセスの形式で公開される。
- (B) 文献情報を数学的・統計的に操作することにより情報伝達および関連の現象や行為を研究する手法。大別すると文献の発生量の計量と利用量の計量があり、分析の対象としては著者、雑誌、索引語、発表年等に関する項目の論文数や被引用数の頻度分布をみることが多い。
- (C) ブログやツイッター、Facebook など、ソーシャルメディアで言及化された回数に基づく研究評価指標。論文の被引用回数等の従来一般的に用いられてきた指標の代替という意味がある。
- (D) 膨大な情報をおさめたデータベースのなかから今まで知られていない知識を発見すること。スーパーマーケットのPOSデータベースから「紙おむつを買う人が一緒に缶ビールを買う傾向が強い」という「知識」が得られたというバスケット分析は、その一例である。
- (E) 平成30年度特許出願技術動向調査報告書 電子ゲーム(特許庁)に含まれる、出願先国別出願件数推移及び出願件数比率(日米欧中韓 ASEAN 各国への出願、出願年(優先権主張年):2010-2016年)を示すグラフである。



- (F) 「自社製品にある機能をつけると、他社の特許に抵触することはないか」など、現在は発生していないが、環境や条件を変えることにより発生する可能性のある問題。
- (G) 「Y社が新しい競合製品を出すといううわさがある。これに対抗するにはどうするか」など、現在はまだ問題は起きていないが、対策をとらない場合は将来的に問題が発生すると予測される問題。
- (H) ブレーン・ストーミングで出された事象を1つずつ書き出し、グルーピングして小さなグループにまとめる。さらにそれを中グループ、大グループに分類していくことで、複数の関連事象を組み立て図解していく。
- (I) 都市の人口滞在の推移や、性別・年齢層や居住者の動向把握など、ビッグデータを用いたデータ分析の事例が見られる。このデータのなかに個人情報が含まれる場合、データの取扱事業者は、個人情報の利用目的を通知または公表することを義務付けられ、利用目的の達成に必要な範囲を超えて、個人情報を取り扱ってはならない。

<解答群>

- |                   |             |               |             |
|-------------------|-------------|---------------|-------------|
| 1. アイゲンファクター      | 2. 自発型問題    | 3. ハゲタカジャーナル  |             |
| 4. インストアマーチャダイジング | 5. データマイニング |               |             |
| 6. 不正競争防止法        | 7. KJ 法     | 8. ポートフォリオマップ |             |
| 9. ビブリオメトリックス     | 10. 環境型問題   | 11. オルトメトリクス  | 12. h-index |
| 13. 設定型問題         | 14. 特性要因図   | 15. ロングテール    | 16. QC7 つ道具 |
| 17. サイトेशनマップ     | 18. 時系列マップ  | 19. 機関リポジトリ   | 20. 個人情報保護法 |
| 21. デジタルアーカイブシステム | 22. 将来型問題   |               |             |

問8 以下はクラウドコンピューティングに関する記述である。空欄 (A) ~ (K) にあてはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

クラウドコンピューティングとは、コンピュータの機能や処理能力、ソフトウェア、データなどを、通信ネットワークを通じてサービスとして呼び出して遠隔から利用する方式のことである。誰でも利用できるようなサービスやシステムを (A)、大企業などが自社ネットワーク上で社員などが利用するために内部的に構築・運用するものを (B)、両者を組み合わせたものを (C) という。

提供する機能や資源の種類により、大きく分けて (D)、(E)、(F) の三つに分類することができる。

(D) は、特定の機能を持つアプリケーションソフトを (G) などを通じて遠隔から操作できるようにしたサービスを指す。利用者はサービスへ登録・加入するだけで、ソフトウェアの入手や導入を行わなくてもすぐに使い始めることができる。データも原則として (H) 側に保管されるため、ソフトウェアやデータの入ったコンピュータを持ち歩かなくても、移動先などで普段とは別の端末からログインして前回の作業の続きを行うことができる。

(E) は、ソフトウェアの実行環境を遠隔から操作できるようにしたサービスを指す。利用者が操作・設定可能なのは (I) よりも上の階層であり、ハードウェアや (J) マシンの動作に直に介入することはできないが、逆に、これらの設定や運用などを自ら行う必要がなく、事業者側にすべて任せられることができると捉えることもできる。

そして、(F) は、(J) 化されたコンピュータ自体を遠隔から操作できるようにしたもので、利用者は (I) などのソフトウェアをすべて自前で用意して導入・運用するものである。メンテナンスや障害発生時などに速やかに別の機材に移転して稼働を続行したり、処理の負荷の増減に合わせて柔軟に資源の追加・削減、つまり (K) ができるといった利点がある。

<解答群>

- |               |               |             |               |                |
|---------------|---------------|-------------|---------------|----------------|
| 1. OS         | 2. ハイブリッドクラウド | 3. Web ブラウザ |               |                |
| 4. プロバイダ      | 5. IaaS       | 6. バージョンアップ | 7. IoT        |                |
| 8. PaaS       | 9. API        | 10. スケーリング  | 11. パブリッククラウド |                |
| 12. PC        | 13. サーバ       | 14. 分散      | 15. 監視        | 16. プライベートクラウド |
| 17. アプリケーション  | 18. SaaS      | 19. ドライバー   | 20. 仮想        |                |
| 21. クラウドストレージ |               |             |               |                |

問9 以下は著作権に関する記述である。空欄(A)～(H)にあてはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

- (1) 著作権法第十五条では「法人その他使用者の発意に基づきその法人等の業務に従事する者が職務上作成する著作物で、その法人等が自己の著作の名義のもとに公表するものの著作者は、その制作の時における契約、勤務規則その他に別段の定めがない限り、(A)」としている。
- (2) 著作物を社会へ広めるのに貢献する演奏家やレコード製作者、放送事業者などを対象に、その実演やレコード、放送などを、(B)で保護している。
- (3) (C)は、書籍、学术论文などの著作物について、適法な複写利用ができるよう、権利者から著作物の複写等の利用に関する管理の委託を受け、集中管理することを目的に設立された。利用者から複写利用料を徴収して著作権者に配分するという集中的権利処理を行い、著作権保護と著作物の適正な利用を実現する。
- (4) 引用の要件は、公表された著作物であること、引用の事実を明らかにすること、引用した部分が明らかであること、引用の目的上(D)であること、複製して引用する場合は出所を明示することである。
- (5) (E)は、米国のように当時方式主義を採用していた方式主義採用国と無方式主義採用国との架け橋となるよう、昭和27年に創設された方式主義をとる条約である。
- (6) 「著作権法及びプログラムの著作物に係る登録の特例に関する法律の一部を改正する法律」が、令和2年6月12日に公布された。具体的には、インターネット上の(F)対策の強化や、写り込みに係る権利制限規定の対象範囲拡大、著作権侵害訴訟における証拠収集手続の強化などが挙げられる。
- (7) AI開発における解析のために著作物を複製する場合や、サイバーセキュリティ確保等のためのソフトウェアの調査解析をする場合に著作物を利用可能とするために、平成30年改正の著作権法30条の4では、著作物に表現された(G)の享受を目的とせず著作物を利用する場合に権利制限規定を設けている。
- (8) 平成30年5月25日公布、平成31年1月一部施行の「著作権法施行規定の一部を改正する法律」で、著作者の許諾が必要な範囲が見直され、より円滑に著作物の利用ができるようになった。具体的には、デジタル化・ネットワーク化の進展に対応した柔軟な権利制限規定の整備、教育の情報化に対応した権利制限規定等の整備、障がい者の情報アクセス機会の充実に係る権利制限規定の整備、(H)に関する権利制限規定の整備が挙げられる。

<解答群>

- |                           |               |             |
|---------------------------|---------------|-------------|
| 1. 実演家人格権                 | 2. 思想又は感情     | 3. その法人等とする |
| 4. その従事する者とする             | 5. 許諾権        | 6. ベルヌ条約    |
| 7. 万国著作権条約                | 8. 出版者著作権管理機構 | 9. 海賊版      |
| 10. アーカイブの利活用促進           | 11. ファクトデータ   | 12. AIの利用   |
| 13. 高齢者の情報アクセス機会の充実       | 14. 正当な範囲     | 15. 50%以内   |
| 16. 30%以内                 | 17. 著作隣接権     | 18. フェイク画像  |
| 19. 日本複製権センター（旧日本複写権センター） |               |             |

問 10 以下は検索エンジンに関する記述である。空欄 (A) ~ (J) にあてはまる最も適切な語句を解答群から選択し、その番号を解答欄に記入しなさい。

1990年代後半のインターネット検索エンジンは、テキストの(A)により検索を行うものが多かった。しかし、検索結果の重みづけの技術が不完全であったため、無関係の Web サイトがタグを悪用し検索結果の上位に表示されるなどの問題が生じていた。

そのような中、1998年に(B)が被リンク数に基づいて検索結果の順位付けを行う技術(C)を開発し、その後同社が大きな影響を与えるようになっていった。

現在、同社のシステムは、複数の(D)で構成され、インデックスに登録されている莫大な web ページの検索を瞬時に可能とし、ユーザの意図に沿った検索結果の表示を行っている。

有用な情報を表示するための要素として用いられるのはページの関連性、出所の専門性、ユーザの(E)等様々なデータが含まれ、急上昇している検索ワードであれば(F)を優先して表示するなど、検索ワードの性質によりその重みづけは調整されている。

同社の検索エンジンには、独自データベースにより検索への回答を直接表示する(G)やユーザの質問に回答すると判断されたページを強調表示する(H)等のツールも組み込まれている。

一方、(B)で満たせない検索ニーズを満たす Web 検索エンジンも存在する。一例としてユーザの個人情報を一切収集しない検索エンジンである(I)や、検索結果の上位を除去することにより埋もれた該当サイトを発見する(J)があげられる。

<解答群>

- |                |              |                   |                   |
|----------------|--------------|-------------------|-------------------|
| 1. 差異          | 2. 類似度       | 3. リンク数           | 4. 検索エンジン         |
| 5. ナレッジグラフ     | 6. Wikipedia | 7. アルゴリズム         | 8. ページランク         |
| 9. SEO         | 10. 適合度の高い情報 | 11. 新しい情報         | 12. 正しい情報         |
| 13. 所属機関       | 14. 位置情報     | 15. Million Short | 16. ScienceDirect |
| 17. DuckDuckgo | 18. Yahoo!   | 19. Google        | 20. Scopus        |
| 21. LexisNexis | 22. 強調スニペット  |                   |                   |

問 1 1 以下はある海外のデータベースの、3つの検索フォームの説明文の抜粋である。  
なお、文中にデータベース名の記載のある箇所は割愛の上抜粋している。

### **General-Purpose Searches** **Companies and Other Filers**

-----

This search works best when you have a specific company in mind. "Search Companies and Filings" allows you to retrieve real-time filings for a specific company and to find key company information — including its name, address, telephone number, state of incorporation, Central Index Key (CIK) number, Standard Industrial Classification (SIC) code, and fiscal year end. Simply type in the name of the company or its CIK number. Note: A CIK is the unique number that the SEC's computer system assigns to individuals and corporations who file disclosure documents with the SEC. All new electronic and paper filers, foreign and domestic, receive a CIK number. You don't need to know a company's CIK, but searching by that number will narrow your search to the exact company you want.

You can also search for all SEC-registered companies in a particular state. Or you can search for all companies with a specific SIC code. For example, if you enter 7370 in the SIC box, you'll retrieve all SEC-registered companies that fall into the Standard Industrial Classification known as "Services-Computer Programming, Data Processing, Etc."

Once you've found the company you want, you can limit the search results by date (for example, "Prior to 2002-01-01") or by the type of filing (for example, "10-K").

### **Latest Filings**

-----

Here you'll find the most recent filings our system has received — from all SEC filers, including individuals and companies, paper filers and electronic filers. The list on this page features the most recent real-time filings from a broad range of companies and individuals for the current official filing date (including filings made after the 5:30 pm deadline on the previous filing day). Note: Filings may be made Monday through Friday (except for U.S. Federal Holidays).

If you want to retrieve a paper filing, please refer to our Fast Answer: Requesting Public Documents. When making your request, be sure to provide the "Paper/Film #" that appears in blue under the company's name.

### **Historical Archives**

-----

Enter a keyword or phrase to search all header information (including addresses) in all filings in the database. You can also use the company CIK as a search term.

Tip: For best results when using the Simple search syntax, use lower-case only and make use of phrase delimiters (commas, "and", "or", etc.), which are described more fully on the bottom of the search screen. For example, to search on "Washington Energy" and "Utility", use the keyphrase washington energy and utility.

(A) このデータベースに収録されている情報として最も適切なものを解答群から選択し、その番号を解答欄に記入しなさい。

- (1) 信用調査報告書
- (2) 有価証券報告書
- (3) 登記簿
- (4) 人物情報
- (5) 国家統計

(B) **Central Index Key Number**(以下C I K)について、最も適切でないものを解答群から選択し、その番号を回答欄に記入しなさい。

- (1) C I Kのみ入力して検索することができる。
- (2) C I Kはコンピュータにより割り当てが行われている。
- (3) C I Kは紙の書類に付与される。
- (4) 特定の検索結果に辿りつくには、C I Kを知らなくてもよい。

以下の文章を読み、説明文から判断しうる範囲で、その検索に最も適切な検索方法を選択し、その番号を回答欄に記入しなさい。

(C) 自動車部品製造業の企業のファイルを検索したい。

- (1) **General-Purpose Searches**
- (2) **Latest Filings**
- (3) **Historical Archives**

(D) 住所から企業のファイルを検索したい。

- (1) **General-Purpose Searches**
- (2) **Latest Filings**
- (3) **Historical Archives**

(E) Apple 社の 2014 年 12 月までの報告書を読みたい。

- (1) **General-Purpose Searches**
- (2) **Latest Filings**
- (3) **Historical Archives**

問12 以下はある文献データベースのヘルプにおける記述である。下線部 (A) ~ (D) の意味について最も近いものを解答群から選択し、その番号を解答欄に記入しなさい。

(A) Computation of similar articles

The neighbors of a document are those documents in the database that are the most similar to it.

(B) The similarity between documents is measured by the words they have in common, with some adjustment for document lengths. To carry out such a program, one must first define what a word is. For us, a word is basically an unbroken string of letters and numerals with at least one letter of the alphabet in it. Words end at hyphens, spaces, new lines, and punctuation. The 132 common, but uninformative, words (also known as stopwords) are eliminated from processing at this stage. Next, a limited amount of stemming of words is done, but no thesaurus is used in processing. Words from the abstract of a document are classified as text words. Words from titles are also classified as text words, but words from titles are added in a second time to give them a small advantage in the local weighting scheme.

(中略)

The local weight of a term is the measure of its importance in a particular document. Generally, the more frequent a term is within a document, the more important it is in representing the content of that document. However, (C) this relationship is saturating, i.e., as the frequency continues to go up, the importance of the word increases less rapidly and finally comes to a finite limit. In addition,

(D) we do not want a longer document to be considered more important just because it is longer; therefore, a length correction is applied. This local weight computation is based on the Poisson distribution and the formula can be found in Lin J and Wilbur WJ.

<解答群>

(A)

1. 類似記事の数え上げ方
2. コンピュータ使用による類似論文作成方法
3. 類似性の評価方法

(B)

1. 類似記事を数えるにはストップワードを除いて行われる。
2. 文書間の類似性は、文書の長さを多少調整した上で、それらが共通して持っている単語によって測定される。
3. 共通の単語数が多ければ多いほど似ていると見なされるため、文書の長さ調整は必須である。

(C)

1. 頻出単語とその単語の重要性の関係には限界がある。
2. 頻度が上昇し続けるにつれ、その用語の重要性は低下する。
3. ストップワードの頻度は飽和状態のため、測定できない。

(D)

1. 長いだけではより重要だと考えられないため、長い文書は必要とされていない。
2. 長い文書が長いというだけでより重要だとみなされてほしくない。
3. ストップワードは長さだけで決定されるわけではない。