

2014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、13：15～14：45 の90 分間です。
4. 中途退席できる時間は、14：00～14：30 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11 問、全12 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 情報検索において、あらかじめ用意された索引語（著者名やキーワードなど）だけでなく、自然文を対象にするようになると、索引語の切り出しが必要になる。

日本語の切り出し方法のうち代表的なものとして、形態素解析とn-グラムがある。それぞれの特徴を比較した表の中の（A）～（F）に当てはまる適切な語句を解答欄に記入しなさい。

	形態素解析	n-グラム
索引語を切り出すための辞書	(A)	(B)
システムのコスト	(C)	(D)
「京都」という語で検索した際にヒットする語	(E)	(F)

<解答群>

1. 安い	2. 「京都」	3. 「京都」「東京」
4. 「京都」「東京都」	5. 高い	6. 設計による
7. 必要	8. 不要	9. 部分的に必要

問2 以下は情報技術に関する記述である。(A)～(E)にあてはまる最も適切な用語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) 組織内のコンピュータネットワークへ外部から侵入されるのを防ぐシステム。また、そのようなシステムが組みこまれたコンピュータ。時には高い性能が要求されるため、専用のハードウェアが用いられる場合もある。
- (B) 文書やデータの意味や構造を記述するためのマークアップ言語の一つである。ユーザが独自のタグを指定できることなどが特徴であり、マークアップ言語を作成するためのメタ言語とも言われる。
- (C) 1991年にフィンランドのヘルシンキ大学の大学院生によって開発された、UNIX互換のOSであり、学術機関を中心に広く普及している。CentOSやUbuntuなどがこれにあたる。
- (D) 拡張子がmovのデータファイルであり、Apple社の開発した動画形式のファイルである。1998年2月には、国際標準化機構(ISO)によって、動画フォーマットの国際標準MPEG-4規格のファイルフォーマットとしても採用された。
- (E) 米国の図書館界で開発されたCJK統合漢字を拡張する形で開発されたものであり、世界中の文字を2バイト(16ビット)で表現したものである。現在Web上で広く普及している。

<解答群>

1. ASCIIコード	2. AVI形式	3. ChromeOS
4. DMZ	5. FirefoxOS	6. firewall
7. HTML	8. JAVA	9. JISコード
10. Linux	11. MacOS	12. QuickTime
13. Shift JISコード	14. WAVE形式	15. XML
16. スイッチングハブ	17. メタデータ	18. ユニコード

問3 以下はデータベース検索の機能に関する説明である。(A)～(I)にあてはまる最も適切な機能、あるいは該当するコマンド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) 索引ファイルに収納されたアルファベット順または五十音順の索引語を列挙し、どんな検索語を使ったらよいか、どこで前方一致をおこなったらよいかを調べることができる機能。
- (B) 語と語の間の距離や語順を指定して検索する手法。
- (C) 前方一致、後方一致等の語幹一致機能。
- (D) ある語が索引されたときにシソーラスや分類の上位語を自動的に付加する機能。
- (E) 一般的に検索結果は時系列順に並べて表示されるが、これを著者名、その他の項目の順に並べ替えて表示させる機能。
- (F) 検索結果中にあるヒットした検索語を太文字や色文字で見やすく表示する機能。
- (G) 文章中から指定したキーワードに加え、その前後の文脈も取り出して表示する機能。
- (H) 複数のファイルで検索を行った時に、ヒットした文献の書誌事項を比較し、これらの中に同一文献であると判定された文献が存在した場合に、優先順位を高く指定した一つのファイルだけで表示させる機能。
- (I) 著者名、発行年などのデータは特別のフィールド(索引項目)に収められており、フィールドを区別して検索を行う方法。

<解答群>

1. EXPAND	2. FIND	3. KWIC
4. LOGOFF	5. SDI (ALERT)	6. SELECT
7. アップosting	8. エンドユーザー検索	9. クイックサーチ
10. コマンド検索	11. シソーラス	12. トランケーション
13. ハイライト	14. ブール演算	15. プルダウンメニュー
16. マニュアル検索	17. 近接演算	18. 検索フィールド指定
19. 重複除去	20. ソート	

問4 以下は問題解決のための代表的な技法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）ブレインストーミング

グループで自由に発言・発想することによりテーマに関する意見を引き出す手法。他者の発言に関して、**(A)** アイデアを出しても良い。多くのアイデアを引き出すことを重視する方法であり、**(B)** を重視する。

（2）ロジックツリー

問題の原因追及などのためにMECE（ミッシー）の考え方に基づいて論理的に**(C)** し、ツリー状に分解・整理する手法。テーマを具体的なレベルまで落としこめるので、対策が立てやすくなる。ただし、事実確認をせずに推論でツリー図を作成する関係上、作成時点での正確さは検証されない。

（3）マトリックス法

比較する二つの要素間の**(D)** を比較することで分析や対策立案に結びつける手法。要素を行と列に分けて配置し、交差する点には各要素の**(D)** やその度合いを記入する。複数の交差表をまとめて取り扱うことにより、問題の全体像の理解や整理が可能となる。

（4）**(E)** 法

ブレイ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、大グループへと分類し、複数の関連事象を組み立てて図解する。その結果、問題解決の対象が明確になり、関連する情報を広く収集できる。日本人考案者のイニシャルより命名された手法。

（5）セブン・クロス法

他の手法などで得られた意見を小カードに書き、それを7項目に大きく分類する。次に、その7つの項目を重要度の高い順に**(F)** へ並べる。さらに、項目毎のカードの重要度を考え、7項目ずつに整理して上から下へ並べる。その結果、タテ・ヨコ7項目の合計49個の一覧表が完成し、多様な意見の全体像把握が可能となる。また、表の左上に向かって重要な項目が配置されることになる。

（6）特性要因図

重要と思われる原因を考えて、図にまとめたもの。書かれた原因は仮説であるため、**(G)** を分析し、調査で確認するための準備の方法である。解決が必要な要因は複数あることが多く、それらを図示することで目的と手段を区別することが可能となり、問題の解決に有効となる。

（7）SWOT分析

組織やプロジェクトの外部環境や内部環境を、強み（Strength）、弱み（Weakness）、機会（Opportunity）、**(H)** の4つの観点で分析する手法である。強みと弱みはその組織の特質であり、他の観点は外部要因である。

<解答群>

1. K J	2. T S	3. 右から左
4. 改良した	5. 階層化	6. 関連性
7. 脅威 (Threat)	8. 左から右	9. 仕事 (Task)
10. 事実と証拠	11. 質より量	12. 主観性より客観性
13. 真の原因	14. 水平化	15. 総合所見 (Total judgment)
16. 討論	17. 否定的な	18. 類似度

問5 以下は情報検索の歴史と仕組みに関する記述である。空欄(A)～(I)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。

情報検索の技術はコンピュータが実用化される前から存在した。これには2つの流れがある。ひとつは図書館の(A)である。近代的な(A)は19世紀から用いられ出しており、1908年に作成された英米コード1908年版が現代の(A)規則の原型である。

もうひとつの流れは学術文献の(B)誌などの二次情報である。19世紀における科学技術雑誌の発展にともない、これらに発表された文献を(B)することが重要となった。現在まで使われている(C)は1879年に、(D)は1907年に創刊されている。

情報検索システムの仕組みは1950-1960年代に開発され、基本的に現在でも使われている。(E)検索は最初にコンピュータによる情報検索が始まった1950年代に、カードや磁気テープに記録された情報を検索するために開発された。データを先頭から順に読んでいき、検索キーワードとの一致を探す方式である。「grep型」とも呼ばれ、オンライン検索が始まる前には広く使われていた。仕組みは簡単だが、対象データ集合が大きくなると(F)がかかる欠点がある。

これに対し、ランダムアクセスの磁気ディスクが利用可能となった1960年代後半に開発されたのが索引検索である。索引検索では、あらかじめデータ集合のレコードから索引語を取り出して、アルファベット順や五十音順の索引ファイルを作っておく方式である。(E)検索のデータベースではレコード番号の下に著者名やキーワードなどの検索データがあるのに対し、索引検索では著者名やキーワードなどの検索語の下にレコード番号のリストを設けた、(G)と呼ばれる仕組みを用いて検索が実行される。検索は50音順やアルファベット順に並べられた索引との一致でおこなわれるので、迅速であり、データが増えても検索速度はあまり変わらない。

さらに上述とはまったく異なる方式として(H)検索がある。検索語と索引語の一致を調べるのではなく、検索文と検索対象レコードの語の(I)の一致度を比較する方式である。

<解答群>

1. Chemical Abstracts	2. EMBASE	3. Index Medicus (MEDLINE)
4. REGISTRY	5. コマンド	6. シソーラス
7. ブール演算	8. マルチファイル	9. 医中誌
10. 概念	11. 近接演算	12. 構造
13. 索引	14. 時間	15. 主題
16. 書誌	17. 抄録	18. 人手
19. 逐次	20. 転置ファイル	21. 頻度分布
22. 目録	23. 要覧	

問6 以下はデータベースに関する記述である。(A)～(L)にあてはまる最も適切な名称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) 米国国立医学図書館(NLM)が作成する世界的な医学文献データベース。
- (B) 米国政府の後援で行われた研究・開発・技術および連邦政府機関やその委託または助成先による分析の概要を収録。州や地方自治体の情報も収録されている。
- (C) 1966年から現在までの海外の教育関係の論文を収録。
- (D) 世界の主要特許発行機関の発行する特許と技術公開誌の情報を収録。電気分野やエンジニアリング分野の図面、および化学構造式、一部の主要国のメインクレームも収録されている。
- (E) 医薬品の有効性、安全性などに関し、国内の医学・薬学関連雑誌の論文情報を収録。
- (F) 物理学、電気工学、エレクトロニクス、コンピュータ分野にわたる世界中の科学技術文献を収録。
- (G) 科学技術全分野に関する文献情報。世界50数か国の情報を収録し、日本語抄録が有用。
- (H) 図書館学および情報科学に関する情報を提供する文献データベース。
- (I) 世界約70か国約7000誌の雑誌から、生物医学および薬学領域の文献を収録。特に医薬品情報が充実している。
- (J) ライフサイエンス分野の代表的文献データベース。特に生物および生物医学分野の広範囲な文献を収集。
- (K) Dun & Bradstreetが提供する世界最大規模の企業情報を収録。世界の2億事業所以上に企業識別コードであるDUNSナンバーを付与。
- (L) 欧州、アジア、日本、ロシア、北米、ヨーロッパの約1億2,000万社の企業財務情報を収録。ビューロー・ヴァン・ダイクが提供。

<解答群>

- | | | |
|---------------------------|------------------|---------------|
| 1. BIOSIS | 2. CiNii | 3. COMPENDEX |
| 4. D&B D-U-N-S Registered | 5. D&B Worldbase | 6. DWPI |
| 7. EMBASE | 8. ERIC | 9. INPADOC |
| 10. INSPEC | 11. JAPICDOC | 12. JMEDPlus |
| 13. JSTPlus | 14. LISA | 15. MEDLINE |
| 16. NTIS | 17. ORBIS | 18. Orbit.com |
| 19. PASCAL | 20. REGISTRY | 21. SciSearch |
| 22. 明日の新薬 | | |

問7 以下は検索システムに関する説明である。(A)～(H)にあてはまる最も適切なシステム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) 帝国データバンク社が提供している、企業情報サービス。信用調査報告書、企業概要・決算書情報、倒産情報などが得られる。
- (B) 1980年に Chemical Abstracts Service によってサービスが開始された、世界の科学技術に関する文献、特許、ファクト情報などを幅広く提供している。
- (C) 400種以上の国内の新聞・雑誌等の記事をはじめ、国内外の企業情報、矢野経済研究所や富士経済の市場調査データ、人物・人事情報、中華圏最大級の中国語と英語の記事データベースである WiseSearch 等、様々なビジネス情報を幅広く収録している。
- (D) NLM から提供されている生物医学分野の文献情報を無料で提供するサービス。約80カ国の5000誌以上の雑誌記事が収録され、古いものは1950年代の文献も含まれている。
- (E) 米国 NASA の情報システムの研究開発において、1960年代に誕生し、1972年から商用システムとして民間での利用が始まった、オンラインデータベースの先駆けである。特許、科学技術、医薬・ライフサイエンス、ニュースなど広範囲のデータベースを提供している。
- (F) LexisNexis 社提供の特許検索データベースで約100カ国のデータが収録されている。
- (G) ジー・サーチ社から提供されているシステムで、100種以上のデータベースが搭載され、企業情報、新聞記事情報、人物プロフィール情報、科学技術情報、法律、特許など様々な分野の情報が集められている。
- (H) 明治以降に発行された特許、実用新案、意匠、商標の公報類、それぞれの出願状況、審査経過情報などが無料で検索できるサイトである。1999年にサービスが開始された。

<解答群>

1. COSMOSNET	2. Espacenet	3. G-Search
4. IPDL	5. NDL-OPAC	6. Patent Scope
7. ProQuest Dialog	8. PubMed	9. SciFinder
10. Scopus	11. STN International	12. TotalPatent
13. Web of Science	14. 医中誌 Web	15. 日経 NEEDS
16. 日経テレコン		

問8 以下は電子ジャーナル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。

電子ジャーナルは商業出版社、大学出版社や学会が作成しており、一般にまとめて検索することは難しい。従って、さまざまなルートで閲覧する方法を心得ておく必要がある。

電子ジャーナルは基本的には出版社によって提供され、出版社毎に独自の検索サイト（プラットフォーム）が構築されていることが多い。主要な商業出版社サイトには、ScienceDirect、SpringerLink、(A)等が挙げられる。主要な学会等でも、独自のサイトの中で電子ジャーナルを提供している場合がある。

また、自身で電子ジャーナル・サイトを構築・運用できない中小出版社のために、電子ジャーナルのホスト・サービス機関が存在し、その主なものとして国内では(B)、海外では HighWire Press、BioOne、PubMed Central 等が挙げられる。

その他、単独では電子ジャーナルの販売・契約が困難な出版社のために、旧来の雑誌取次業者などが電子ジャーナルの販売・契約業務を代行している場合があり、これらを(C)と呼ぶ。

学術雑誌は電子ジャーナル購読を軸とする利用形態に移行しつつあり、相対的に冊子体での出版は縮小、あるいは廃止の傾向にある。電子ジャーナルの普及は、同一機関内における重複購読を解消し得る反面、出版社側の価格設定にも大きな影響を与えており、(D)とも呼ばれる包括契約や、機関規模に応じた価格帯設定が採られる場合がある。

近年の学術雑誌価格の高騰を受けて、学術論文の無料公開に関する(E)運動が欧米で高まっている。また、大学の図書館等では、所属研究者の研究論文その他のデジタルコンテンツを集めて公開する(F)が広がりを見せている。

<解答群>

1. J-Science	2. J-STAGE	3. MatNavi
4. Webcat Plus	5. Wiley Online Library	6. アクセス・オン・デマンド
7. アグリゲータ	8. インクルード	9. オープンアクセス
10. オープンリポジトリ	11. ディストリビュータ	12. ディスパッチャー
13. ビッグ・ディール	14. フリーフォーマット	15. プロバイダ
16. 機関ダイレクト	17. 機関リポジトリ	18. 電子図書館

問9 以下は検索結果の評価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。

<再現率と適合率>

ブール演算子を用いる伝統的な検索システムでは検索結果の評価に「再現率」と「適合率」を用いることができる。

「再現率」とは検索対象集合の中にある〔A〕をどのくらい取り出すことができたか、という比率である。「適合率」とはヒットした回答集合の中に〔A〕がどのくらい入っていたか、という比率である。「再現率」と「適合率」が示す指標および向上させる方法を下記に示す。

	再現率	適合率
指標	〔B〕	ノイズ
向上させる方法	<ul style="list-style-type: none"> ・ 〔C〕または関連語の追加 ・ 語幹一致検索機能の活用 ・ 〔D〕用語の活用 	<ul style="list-style-type: none"> ・ 〔E〕のための検索語を追加 ・ 検索を特定の〔F〕に限定 ・ ノイズの原因となる検索語を除去

一般に再現率と適合率は〔G〕のような関係にあるといわれる。再現率を上げようとするともノイズが増えて適合率が下がり、適合率を上げようすると〔B〕が生じて再現率が下がる。

再現率と適合率を両方あげることがサーチャーの腕の見せどころである。

<ランキング>

検索エンジンでは検索回答数（回答見積数）は数万から数百万にのぼり、全件を調べることは実用的ではない。そこで検索要求に合致していると思われる順にヒットしたページを並べる機能をランキングという。ランキングには次のような条件を考慮する。

- a. その検索語のページ内での〔H〕
- b. 複数の検索語がある場合はそれらの〔I〕
- c. ページの重要度

ページの重要度は次のような条件で判断する。

- a. ページ製作者の重要度（公的な機関であるか、など）
- b. ページの人気（他のページからのリンクが多いページは点数が高い）
- c. ページの新しさ
- d. ページの〔J〕数

<解答群>

1. アクセス	2. インデックス	3. コード
4. コマンド	5. シソーラス	6. センテンス
7. ソート	8. ヒットターム	9. フィールド
10. リポジトリ	11. 拡張	12. 近接度
13. 検索漏れ	14. 絞り込み	15. 辞書
16. 出現頻度	17. 全レコード	18. 適合レコード
19. 同義語	20. 反比例	21. 比例
22. 密接	23. 類似度	

問10 以下は情報源及び情報流通に関する記述である。(A)～(H)にあてはまる最も適切な名称を解答群から選び、その番号を解答欄に記入しなさい。

- (A) 『国史大事典』『古事類苑』『東洋文庫』などを含む百科事典のWeb版である。
- (B) 国公立の各大学から送付される印刷版の学位論文(博士)を所蔵しており、所蔵資料の大規模デジタル化事業の一環として、著作権者の許諾が得られた博士論文をインターネット上で公開するサービス。
- (C) 全国の大学図書館等によるオンライン共同分担目録方式の所蔵資料の総合目録データベースであり、それを利用した図書館間の相互貸借等を行うサービスもあわせて提供している。
- (D) 書誌ユーティリティとして、納本制度に基づく「日本全国書誌」をインターネット上で提供し、集中目録作業方式により作成した機械可読目録「JAPAN MARC」を各機関に頒布している機関の名称。
- (E) 登録研究者同士のディスカッション・情報共有等を実現する研究コミュニティなども充実した、循環型の情報活用基盤サービスである。国内最大の研究者情報を収録した、日本の研究者総覧データベースである。
- (F) 電子リソースから収集した学術情報や研究資料について、管理や共有、情報発信ができるウェブサービス。
- (G) 図書資料などについて、キーワードだけでなく自然文による検索が可能であり、図書の内容や目次の情報も得られる。NIIがサービスを提供している。
- (H) 政府統計データの総合的な窓口であり、各省庁の様々な統計データを取得できる。

<解答群>

1. AMS Books Online	2. bookplus	3. CiNii Books
4. EBSCO	5. e-Stat	6. Google Books
7. integrum	8. J-STAGE	9. JSTOR
10. magazineplus	11. NACSIS-CAT	12. NDL
13. RefWorks	14. researchmap	15. Webcat Plus
16. Web Oya-Bunko	17. ジャパンナレッジ	18. ブリタニカ・オンライン・ジャパン
19. 官報情報検索サービス	20. 国立国会図書館デジタルコレクション	
21. 国立国会図書館リサーチナビ		

問 1 1 以下の英文は、Elsevier 社への論文投稿に関する用語定義の一部を示したものである。これを読み、対応する日本語の空欄 (A) ~ (I) にあてはまる最も適切な語句または文を解答群から選択し、その番号を解答欄に記入しなさい。なお、日本語の一部の表現については、意識とした箇所もあるので注意すること。

① Pre-print Definition:

A pre-print is an author's own write-up of research results and analysis that has not been peer-reviewed, nor had any other value added to it by a publisher (such as formatting, copy editing, technical enhancement, etc.).

① プレプリントの定義:

プレプリントとは、(A) 研究結果と分析であり、(B) や出版社による付加価値 (書式、コピーエディティング、技術的な改善など) は加えられていません。

② Accepted Author Manuscript (AAM) Definition:

An accepted author manuscript (AAM) is the author's version of the manuscript of an article that has been accepted for publication and which may include any author-incorporated changes suggested through the processes of submission, peer review and editor-author communications. AAMs do not include other publisher value-added contributions such as copy-editing, formatting, technical enhancements and pagination.

② アクセプトされた著者原稿 (AAM) の定義:

アクセプトされた著者原稿 (AAM) とは、出版するために受理された論文原稿の著者版であり、投稿、(B)、編集者とのやり取りを通じて (C) が含まれます。AAM には、コピーエディティング、書式、技術的な改善、(D) など、出版社による付加価値を (E)。

③ Published Journal Article (PJA) Definition:

A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publisher activities including copy-editing, formatting and pagination.

③ (F) (PJA) の定義:

(F) (PJA) とは、ジャーナルに掲載された、あるいは掲載予定の論文の最終版です。コピーエディティング、書式、(D) など、付加価値を加えるための出版社によるすべての作業を終えています。

④ Commercial use/purpose means:

the use or posting for commercial gain, whether for a fee or for free.

④ (G) とは:

(H)、営利目的で論文を使用または (I) することを意味します。

<解答群>

1. ページ番号の付与	2. 加えています	3. 加えていません
4. 監査	5. 掲載	6. 査読
7. 出版済みジャーナル論文	8. 商業利用	9. 精読
10. 装丁	11. 著者が加えた変更	12. 著者が得た原稿料
13. 著者の同意	14. 著者自身が執筆した	15. 著者自身により公になった
16. 翻訳業務の委託	17. 有料、無料を問わず	18. 料金を受け取り