### 02139 计算机信息检索

### ---------- 名词解释题 ----------

1. 网络资源目录

***网络目录是以整个互联网资源为检索对象，将一些重要的、优质的网络资源组织成分类目录，提供给用户查询检索***

1. FTP

***FTP是因特网使用的文件传输协议。该协议的主要功能：完成从一个系统到另一个系统完整的文件拷贝，即在因特网的联网计算机之间传输文件***

1. 词典
2. 网页快照
3. 政府信息

***政府信息是指国际组织、各国政府及其相关部门所发布的信息***

1. 布尔检索

***用户根据所检索关键词在检索结果中的逻辑关系递交检索，检索模块根据布尔逻辑的基本运算法给出检索结果***

1. 元搜索引擎

***元搜索引擎就是通过一个统一的用户界面，帮助用户在多个搜索引擎中选择和利用适用的搜索引擎来实现检索操作***

1. 浏览器

***浏览器是指Web服务客户端的浏览程序，是显示网页服务器或档案系统内的文件，并让用户与这些文件互动的一种软件。***

1. 关键词

***用以描述资源主要内容的关键词语或分类号码表示的有代表性的主题词***

1. 学科分类法
2. 基于内容的图像检索（CBIR技术）
3. 参考信息

***参考信息是指人名、地名、机构、事件、统一数据等一类数据、事实信息，是人们在工作研究和日常生活中经常查考、引用的信息***

1. URL

***URL是通用资源定位程序又称网络资源的统一定位格式或统一资源定位器，实际上就是一个用以标识文档类型及其所在网络地址的字符串。***

1. 信息需求

***“需求”可以是个人某个不确定的事物，而“信息”则是可以减少这种不确定性的因素，所以“信息需求”是指个人内在的认知状态，与外在环境接触后所产生了不确定性，发现了个人认知的缺口，进而试图寻找适当的信息，以弥补认知缺口，解除上述不确定性***

1. 超文本标记语言(HTML)

***html是一种专门编程语言，具体规定和描述了文件显示的具体格式。***

1. 学科信息门户
2. 白页信息
3. 基于文本的图像检索(TBIR)
4. 超文本传输协议HTTP

***超文本传输协议http是浏览器与web服务器之间相互通信、传输、响应用户请求的协议。***

1. 搜索引擎

***搜索引擎是一种能够通过互联网接受用户查询指令，帮助用户迅速的从网上查找所需要的信息，并向用户提供符合其查询要求的搜索结果列表以及相关信息的检索系统。***

1. 学位论文
2. 链接

***链接是web网页的元素，是指向其他资源的指针，把web页捆绑在一起，跟随链接可从一个文档跳转到另外一个相关文档。***

1. 搜索工具门户
2. 图像检索
3. 黄页信息
4. 会议文献
5. 年鉴

***年鉴是指每年出版的统计数据和事实的汇编，兼有现期性和回溯性内容。其包含的地区和学科领域一般比较广泛，也可能限于某一特定国家或特定的领域***

### ---------- 判断改错题 ----------

1. 上位词是指概念上内涵更窄的关键词 【 更广 】
2. OCLC是联机计算机图书馆服务的研究机构 【】
3. 白页提供对机构信息的检索 【 人物信息 】
4. 在Google的检索框中的多个关键词间的缺省逻辑运算是或运算 【】
5. HTML协议是因特网上使用的文件传输协议 【 FTP协议 】
6. Google搜索引擎支持不完全的布尔逻辑 【】
7. 并行式元搜索引擎拥有独立的网络资源采集标引机制和相应的数据库 【】
8. Google中的关键词间的逻辑或运算符是or 【】
9. Medline是法律文献资源的检索工具 【】
10. FTP信息资源是互联网上最为流行的信息传播方式 【 WWW信息资源 】
11. 元数据是用以揭示各类电子文档的内容和其他特征的有力工具 【 √ 】
12. 网络信息检索技术不断发展，基于文本的多媒体检索更为业界关注 【】
13. Google实现布尔或运算可以使用OR或or 【】
14. 长颈鹿是动物的上位词 【 下位词 】
15. 依据资源的网络传输协议可将其分为Web资源、Telnet资源及FTP资源等 【 √ 】
16. 垂直搜索引擎的信息量多于通用搜索引擎的信息量 【 少于 】
17. 第二代搜索引擎是以文档分类导航为特征且以Yahoo为代表 【】
18. 黄页提供著名人物生平事迹的检索 【 白页 】
19. 搜索引擎不会侵犯公众隐私 【】
20. 检索效率是评价与选择因特网信息检索工具的核心指标 【 检索功能 】
21. 最有代表性的索引型检索工具就是搜索引擎 【 √ 】
22. 小麦是粮食的下位词 【 √ 】
23. Google搜索引擎支持不完全的布尔逻辑 【 √ 】
24. 互联网信息检索时，搜索引擎仍然不能找到大量“隐形的网页” 【 √ 】
25. “中国优秀博士学位论文全文数据库”是属于万方数据和知识服务平台的资源 【】
26. 基于文本的图像检索技术TBIR更符合用户的使用习惯 【 √ 】
27. 垂直搜索引擎对信息采集的广度要求高于深度要求 【 低于 】
28. 百度百科旨在创造一个涵盖各领域知识的中文信息收集平台 【 √ 】
29. 政府信息可以说是网上最具有价值的信息，由于出自官方，所以具有权威、可靠等特点 【 √ 】
30. Telnet协议是互联网远程登录服务的标准协议和主要方式 【 √ 】
31. Google被称为是全球第一大搜索引擎 【 √ 】
32. OCLC是一个面向用户的交互式的期刊文献联机检索系统 【】
33. 互联网信息检索时，搜索引擎是网络检索工具的典型代表 【 √ 】
34. 中国高等教育文献保障系统（CALIS）是商业性学位论文数据库 【】
35. 百度图片是基于文本的图像搜索引擎 【】
36. 通用搜索引擎进行检索的最小单位是结构化的数据 【】
37. 不列颠百科全书网站是以大英百科全书为主要资源的综合性百科知识检索网站 【】
38. 互联网信息检索具有近似无限的检索范围，动态的检索对象 【 √ 】
39. 万方数据知识服务平台的学位论文库主要收录美欧的学位论文 【】
40. 目前可以用来处理图像的主要特征包括颜色、文理和形状 【】
41. 通用搜索引擎具有专、精、深的特点，具有行业色彩 【 垂直搜索引擎 】
42. 维基百科是一个语言、内容开放的网络百科全书计划 【】
43. 按照一定的评价标准对被评估站点的各方面特征、质量做出主观评判的评价方法叫做定量评价 【】
44. 较高的检准率可以节约用户从数据库中查找相关信息的时间 【 检全率 】
45. 在独立网站上运行，以互联网信息内容为检索对象的常规检索工具叫做桌面搜索工具 【】
46. 拟声查询是指用户发出与要查找的声音相似的声音来表达检索要求 【】
47. MediaFinder是美国和加拿大最大的连续出版物和目录数据库 【】

### ---------- 简答题 ----------

1. 简述网上参考信息资源的主要类型
2. ***百科知识检索网站***
3. ***人物信息检索网站***
4. ***地理信息检索网站***
5. ***时事、新闻信息检索网站***
6. ***机构信息检索网站***
7. ***语词信息检索网站***
8. ***统计信息检索网站***
9. 简述因特网上可找到信息的类型
10. 简述可用于帮助图像分析、标引的Web文档内容
11. 简述WEB信息资源检索工具的类型
12. 简述社会科学信息检索工具SOSIG的功能组成部分
13. 简述Uncover检索系统及其主要联机服务
14. 简述中国期刊网检索工具包含的数据库
15. 简述用户服务组的功能
16. ***订阅该组***
17. ***取消订阅该组***
18. ***读取组中文章***
19. ***发布文章至该组***
20. ***跟随文章***
21. ***回复***
22. ***转发***
23. 简述使用基于关键词检索网络信息资源的优点
24. 列举与WWW信息资源相关的重要概念
25. ***超文本传输协议HTTP***
26. ***通用资源资源定位程序URL***
27. ***超文本标记语言HTML***
28. ***主页***
29. ***链接***
30. ***浏览器***
31. 简述搜索引擎工作原理
32. ***从互联网上抓取网页***
33. ***建立索引数据库***
34. ***在索引数据库中搜索排序***
35. ***对处理结果进行处理排序***
36. 简述有助于基于文本的图像检索方法的Web文档内容
37. 简述网络信息检索技术发展方向
38. 简述互联网信息检索的类型
39. ***“顺链而行”的浏览***
40. ***基于目录型检索工具的互联网信息检索***
41. ***基于索引型检索工具的互联网信息检索***
42. 简述搜索引擎信息预处理技术
43. ***关键词提取技术***
44. ***垃圾信息消除技术***
45. ***链接分析技术***
46. ***网页重要性计算技术***
47. 简述目录型网络检索工具的特点
48. ***面向用户的易用性***
49. ***提供准确性较高的信息内容***
50. ***提供导航作用良好的检索体系结构***
51. ***以明确的主题和对象为中心设置确定类目***
52. ***有兼容性强的多元类目***
53. 列举搜索引擎的检索功能

***一般检索功能：***

1. ***布尔逻辑检索***
2. ***词组检索***
3. ***截词检索***
4. ***字段检索***
5. ***限制检索***

***特殊检索功能：***

1. ***自然语言检索***
2. ***多语种检索***
3. ***区分大小写的检索***
4. 简述对因特网信息资源进行评价的意义

***面对网络信息资源的质量不均衡状态，发展网络信息资源的评价方法及相关标准和工具是至关重要的。对网络信息资源进行质量评价，从信息海洋中经过甄别，筛选出有学术价值或利用价值的精华部分，推荐给网络用户使用，可以较好的屏蔽一些信息污染或检索噪音，大大地提高用户利用互联网信息资源的效率***

1. 简述目录型网络检索工具的特点
2. ***面向用户的易用性***
3. ***提供准确性较高的信息内容***
4. ***提供导航作用良好的检索体系结构***
5. ***以明确的主题和对象为中心设置确定类目***
6. ***有兼容性强的多元类目***
7. 简述目录型网络检索工具的使用方法
8. ***采用广度优先查询法***
9. ***注意了解各个检索工具的收录特点，选择适当的工具***
10. ***掌握各目录型检索工具的分类体系***
11. ***关注热门主题的常用目录***
12. ***关注开放目录型检索工具***
13. 学术期刊有什么特点？在网联网上检索学术期刊的途径有哪些？

***特点：品种多、数量大、涉及面广、拥有庞大的写作队伍和广泛的读者群***

***途径：访问著名的出版公司、学术团体的网站***

1. 作为数字化、网络化信息的核心集成，互联网上的信息内容包括哪几种？其中隐形的网页数量多吗？
2. 通用搜索引擎的优势主要有哪些？请列举有代表性的通用搜索引擎，国内外各两个

***优势：全文搜索、检索功能较强、信息更新快***

***国外：Google、Bing***

***国内：百度、搜狗***

1. 目录型网络检索工具有什么样的结构特点？使用时最好注意哪些方法
2. 简述通过FTP获取文件的一般步骤
3. ***登录***
4. ***浏览目录，找到所需文件***
5. ***设置文件传输参数，选择文件的传输模式***
6. ***下载所需文件***
7. 简述因特网信息资源的一般评价方法
8. ***从URL中得到信息***
9. ***查看网站描述和版权说明***
10. ***从链接中获取信息***
11. ***参考他人评价***
12. 简述因特网信息检索的主要特点
13. 简述元搜索引擎与普通搜索引擎的区别
14. ***元搜索引擎可以将一次提问提交多个数据库***
15. ***元搜索引擎没有自己独立的资源库***
16. ***元搜索引擎的结果是基于独立搜索引擎结果的二次加工***
17. ***元搜索引擎标明结果记录的来源搜索引擎及其相关度***
18. 网络信息检索技术发展需要以什么为核心？主要在哪6个领域进行开发和创新？
19. 网络信息资源质量的评价标准主要分为哪5条？评价方法主要分为哪两类？
20. 网络检索工具、按检索对象、检索机制和集成检索工具数量的不同，分别如何进行划分？
21. 在信息检索策略中，试给出任意7个进阶检索技巧
22. 什么是用户组信息资源，它的主要优点是什么？

***用户组信息资源是由一组对某一特定主题事物有共同兴趣的网络用户组成的邮件群组或电子论坛，是一种有组织的电子邮件系统，用户在其间通过电子邮件进行一对多的交流。***

***优点：***

1. ***直接交互***
2. ***全球互联***
3. ***主题鲜明***
4. 互联网检索工具的性能评价主要从哪5个方面进行？其中反映检索系统效率的主要指标是什么？
5. ***收录范围***
6. ***检索功能***
7. ***检索效率***
8. ***检索结果的处理和展示***
9. ***用户界面设计***

***反映检索系统效率的主要指标是检全率与检准率。***

1. 什么是垂直搜索引擎？主要特征是什么？其关键技术主要有哪两项？

***垂直搜索引擎是针对某一特定领域、某一特定人群或某一特定需求提供的信息搜索和相关服务。***

***主要特征是专、精、深，且具有行业色彩***

***关键技术主要有“主题爬虫”技术和web信息提取技术***

1. 目录型网络检索工具主要适用于什么检索情况？请列举有代表性的网络资源目录，国内外各两个。
2. 简述互联网信息资源的局限性
3. ***分散***
4. ***无序***
5. ***多变***
6. ***信息质量难以控制***
7. ***信息安全难以保障***
8. 简述搜索引擎的主要工作步骤
9. 简述图书馆馆藏目录检索系统的检索途径

### ---------- 论述题 ----------

1. 试述影响选择及使用元搜索引擎的因素
2. 试述影响搜索引擎的检索结果相关性的各种因素
3. 试述检索策略的制定步骤
4. 试述目录型网络检索工具的使用方法
5. 试述常用的检索技巧
6. 试述网络信息检索中的法律问题
7. 试述网络信息检索工具的性能评价指标
8. ***收录范围：每种互联网信息检索系统都有特定的收录对象与收录原则、方针，选择检索系统必须要了解其收录数据的性能。***
9. ***检索功能：包括检索方式单一还是多样；所使用的技术是否先进、多样；对所查信息是否有选择与限定的自由。***
10. ***检索效率：较高的检全率可以节省用户从数据库中查找相关信息的时间。较高的检准率可以节省用户从检出的所有文档中过滤掉无关信息的时间。信息检索的目标是满足高检全率与高检准率，但两者难以同时兼顾。当要求系统提高检全率时，检准率会下降，反之亦然。***
11. ***检索结果的处理和展示：如果系统呈现检索结果的方式不佳则会阻碍用户浏览和吸收信息的效果。***
12. ***用户界面设计：若设计不良，即使检索功能再丰富、再先进，也可能无法吸引用户的使用。***
13. 试述网络信息检索工具的分类及其具体类型

***网络信息检索工具的分类可以从多个角度和标准来进行。一般可分为以下两大类：***

***web资源检索工具：以web资源为检索对象，又以web形式提供的检索工具，代表网络信息检索的较高水平，且应用较为普遍。包括：***

1. ***目录型检索工具***
2. ***搜索引擎***
3. ***多元搜索引擎***

***非web资源检索工具：以非web资源（如FTP，Gopher等）为检索对象的检索工具：包括：***

1. ***Archie***
2. ***Hytelnet***
3. ***Deja News***
4. ***Medline数据库***
5. ***ERIC数据库***
6. 试述检索策略的制定步骤及其实施步骤

***制定步骤：***

1. ***分析检索课题，明确信息需求。要想让互联网信息成为自己的知识库，并不是要占有全部的互联网信息，而是能够在互联网上找到自己需要的信息，所以正确的对信息需求进行描述，是互联网检索的第一步***
2. ***选取关键词。关键词的正确选取，可以有效的提高检索的效率***
3. ***关键词的组配。一个关键词只能检索出目标信息的一部分或某个知识点，一般不能单独满足信息需求的完整内容。我们要根据具体的信息需求来设计和编制检索表达式，也就是将多个关键词进行某种组合，以便系统、全面地检索出所需要的信息***
4. ***选择检索工具，实施检索策略***

***实施步骤：***

1. ***判断目标信息可能存在的地方，对目标信息进行推测和判断，建立在对大量信息资源非常熟悉的基础上， 我们可发根据自己的经验作出检索判断***
2. ***了解互联网信息检索的特殊性，在使用互联网进行检索时需要清楚，所谓的“互联网信息检索”并不是直接在万维网上寻找某个文档***
3. ***了解可用的检索工具。如：搜索引擎、元搜索引擎、网络资源目录和专门的数据库***
4. ***仔细分析自己的信息需求并选择合适的工具。需求决定了检索结果，在检索之前对检索主题进行仔细分析，考虑清楚自己要找的是什么，再确实用什么工具进行检索。***
5. ***根据检索结果调整检索策略，上一步的检索未能得到让自己完全满意的信息，那就很可能是检索提问还有改进的地方***
6. 书籍是重要的学习资源，在互联网上检索书籍的途径有哪些？ 各有什么特点？
7. 目录型检索工具的特点有哪些？其分类体系常用哪些分类法？它如何引导用户在目录型检索工具中操作以满足查询需求？
8. 试述网络信息检索工具的工作原理
9. ***网络信息检索工具的工作原理主要是指基于索引型检索工具的工作原理***
10. ***主要分为信息组织和信息查询两个环节，对于互联网信息来说，这两个部分相辅相成。***
11. ***信息组织：1.文档分析：计算所能看到的知识一堆文字或图像，对其内容无法进行识别。如果要让计算机处理万维网中信息，就必须首先将这些信息加工成计算机可以理解的信息形式后才能进行处理。文档分析的主要功能是过滤文件系统信息，为文件系统的表达提供一种满意的索引输出。2.建立索引：为了快速响应检索，在组织信息时就必须建立索引。索引是一种将关键词词目映射到相关文档的数据结构。3归类和聚类：自动分类方法。4.元数据：制定一套数据描述格式来有效率地描述网上的数字化信息资源***
12. ***信息查询：包括布尔检索模型：用户根据所检索关键词在检索结果中的逻辑关系递交检索，检索模块根据布尔逻辑的基本运算法给出检索结果、扩展布尔模型：对布尔逻辑模型的改进、向量空间模型：文档和检索提问都可以用向量表示、概率检索模型：在布尔逻辑模型的基础上为极倔检索中存在的一些不确定性而引入的。***
13. 试述元搜索引擎的优缺点

（1）信息覆盖面打，检索效率高。元搜索引擎集成多个成员搜索引擎，通过并发查询、增加检索范围、扩大查询区域、信息覆盖率增加，以获得较高查全率。

1. 搜索引擎检索的对象是实时匹配互联网上的网页吗？说明其工作原理的四个步骤，以及工作需要的四种主要信息预处理技术。
2. 在使用搜索引擎时，关键词的有效选取非常重要；请说明关键词选取常用的5步骤，以及常见的可配套使用的布尔逻辑操作符。
3. 常用的互联网信息检索模型有哪些？各自的优缺点是什么？
4. ***布尔检索模型：优点是原理简单易理解，容易在计算机上实现并且检索速度快。缺点是最终给出的检索结果没有相关性，不够精准，不能反映不同的索引项对一个文档的重要程度的差异***
5. ***扩展布尔模型：对布尔逻辑模型的改进，将非此即彼的判断方式改为计算相关度***
6. ***向量空间模型：优点在于使用便捷，并且在模型中有许多可调整的计算方式。缺点在于缺乏理论的支持和验证，***
7. ***概率检索模型：优点在于有严格的数学理论基础，采用了相关反馈原理克服不确定性的推理。缺点是参数估计的难度比较大，文件和检索的表达也比较困难。***
8. 什么是元搜索引擎？它如何工作？它与普通的搜索引擎主要的区别是什么？
9. 试述目录型网络检索工具与搜索引擎的区别

从两者的适用性、优势与不足以及发展趋势等角度来进行比较：

1. ***适用性：目录型网络检索工具将信息系统分门归类，把同一主题的网站信息组织在一起并按一定顺序排列，通过主题分类层层浏览，比较符合人们传统的信息查找方式，尤其适合那些希望了解某一方面、某一概念的全面信息，而不是只限于查询若干关键词的用户，有助于逐步缩小主题或者查找某个主题常见的、质量较高的信息。特别适用于一般的网络信息门户，面向普通用户及网络新手展示、提供信息时采用。主要适用于：用户进行较笼统的主题浏览和检索。当用户尚未形成很精确的检索概念时，采用主题指南作为检索起点非常有效。***
2. ***优势与不足：目录型网络检索工具由人工编制维护，在信息收集、编排、编码及摘要简介等方面，需要投入大量人力和时间，它强调是组织、浏览功能。其优点是：收录的网站经过专家的人工精选，故内容丰富、学术性强，信息质量较为稳定。人工干预提高了其返回结果的相关性，提高了检索的查准率。信息经过系统归类，分类浏览方式直观易用，适合多数网络用户新手。由于目录是依靠人工来评价的，所以从目录搜索到的结果更具有参考价值。缺点：由于受人工标引的限制，其搜索范围较搜索引擎要小许多，信息更新速度也相对较慢。保证目录结构的清晰性与每一类目下的条目数量也形成了矛盾，因此极大地限制了网络信息资源的内容。收录的网页数量和标引深度相对不足，很难检索到较专业、深入的信息。难于控制主题等级类目的质量，各系统的分类体系和标引方式不统一，不规范，导致一定的误查和漏查。搜索引擎的优点：查询全面而充分、方便快捷。缺点是可供选择的资源太多，基于关键词检索的搜索引擎信息组织与标引缺乏控制，没有从概念语义上匹配，没有范畴限定，信息查询的命中率、准确率、查全率不太令人满意，信息噪音太大。由于没有分类目录那样清晰的层次结构，令用户感觉繁杂。同时，信息资源缺乏规范控制，降低了可靠性。需要有明确的检索词，并具备一定的检索知识，了解逻辑组配语法，对缺乏检索知识和技能的用户有较大难度。***
3. ***目录型网络检索工具与搜索引擎的发展：目录型网络检索工具和搜索引擎都是对网络哦信息资源进行组织和揭示的工具，但是不同性质的两种检索系统。面对汹涌的网络资源和用户，理想的检索工具是能在目录型网络检索工具主题分类结构下实现搜索引擎的全文检索，将两者的优势结合起来，使检索结果更为全面、准确。***
4. 试述搜索引擎的“瓶颈”