

电子商务复习

电子商务的概念

电子商务(Electronic Commerce, 简称 EC), 它是指在全球各地广泛的商业贸易活动中, 在因特网开放的网络环境下, 基于浏览器/服务器应用方式, 买卖双方不谋面地进行各种商贸活动, 实现消费者的网上购物、商户之间的网上交易和在线电子支付以及各种商务活动、交易活动、金融活动和相关的综合服务活动的一种商业运营模式。

PS:如今的电子商务不仅仅是基于浏览器/服务器应用方式。

什么是电子商务?

电子是手段

商务是目的

广义: 使用各种电子工具从事商务活动

狭义: 主要利用Internet从事商务活动

电子商务的类型

电子商务主要类型包括八类, 分别是: B2B、B2C、C2C、C2B、 B2M(Marketing)、M(Manufactures)2C、B2A (即B2G) 、C2A (即C2G) 。

2003年12月, 慧聪网(<http://www.hc360.com/>)实现了在香港联交所创业板的成功上市, 成为国内B2B电子商务领域的首家上市公司

2007年11月6日, 电子商务巨头阿里巴巴(<http://www.alibaba.com/>)在香港上市

2010年10月26日, 麦考林(<http://www.m18.com>)成功登陆纳斯达克, 成为中国B2C第一股

B2B(Business to Business)

商家(泛指企业)对商家的电子商务, 即企业与企业之间通过互联网进行产品、服务及信息的交换。

通俗的说法是指进行电子商务交易的供需双方都是商家(或企业、公司), 它们使用了Internet的技术或各种商务网络平台, 完成商务交易的过程。这些过程包括: 发布供求信息, 订货及确认订货, 支付过程及票据的签发、传送和接收, 确定配送方案并监控配送过程等。

B2C(Business to Customer)

B2C模式是我国最早产生的电子商务模式，以 8848 网上商城正式运营为标志。

第一种：**综合商城**

将现实中的商城搬到网络上，与现实生活中的商城相对应。城中有很多店。天河城、贝斯尼尔、正佳广场等。

第二种：**专一整合型**

赛V网：赛V网主要从事体育用品网上销售、导购、新闻资讯、赛事报道等，将作为一体的大型体育综合门户网

第三种：**百货商店**

商店，谓之店，说明卖家只有一个；而百货，即是满足日常消费需求的丰富产品线。这种商店是有自有仓库，会库存系列产品，以备更快的物流配送和客户服务。

亚马逊、京东网、当当网、苏宁易购等

第四种：**垂直商店**

服务于某些特定的人群或某种特定的需求，提供有关这个领域或需求的全面产品及更专业的服务体现。

红孩子、新蛋网、尊享网等。

第五种：**复合品牌店**

随着电子商务的成熟，将有越来越多的传统品牌商加入电商战场，以抢占新市场，拓充新渠道，优化产品与渠道资源为目标。佐丹奴、百丽等。

第六种：**轻型品牌店**

小品牌也有大作为。

VANCL（凡客诚品）等。

第七种：**服务型网店**

满足人们不同的个性需求。

易美（网上冲印公司）、亦得（亦得代购 买遍世界）等。

第八种：**导购引擎型**

作为B2C的上游商，给商家们带去客户。

爱比网（比友们可以通过这里分享到比友的产品体验点评，比友们也热衷于将自己用过的产品体验告诉给更多的比友。）等。

C2C(Consumer to Consumer)

C2C 是**用户对用户**的模式，C2C 商务平台就是通过为买卖双方提供一个在线交易平台，使卖方可以主动提供商品上网拍卖，而买方可以自行选择商品进行竞价。C2C的典型有淘宝网、拍拍网等。

C2B(Consumer to Business)

C2B 模式指的是**消费者对企业**，这是电子商务领域新出现的模式。

C2B 模式是先在网聚合一个庞大用户群，形成一个社区，再以团购等形式，用户获得批发商的价格。

国内做的比较好的 C2B 网站有中国团购在线、一起买等。

电子商务的商业模式及优缺点

店面模式

✂ 一个真实的电子商务站点，以某种价格提供产品或货物。它是一种典型的提供实际商品货物和服务的方式，比如书籍、数码产品或一项鲜花送货服务。商家直接把货发送给客户，销售中没有零售商或中介

网络加传统模式

✂ 网络加传统的商店把网站和传统店面结合起来。它的优势是已经建立起了一个品牌，可以用传统店面去宣传它的网站。

按订单生产的商家模式

✂ 制造商可以采用这种模式提供商品或服务，也有能力接收定制的订单。定制产品然后个性化地生产并送货给客户，为客户提供了增值服务，并使得制造商只需制造那些能卖得掉的产品。

✂ DELL 的故事……

服务提供商模式

✂ 提供相关的咨询服务并需要用户支付一定的费用，如在线资料下载服务，还有一些服务提供商提供基于广告的接入服务，通过从广告商那里获取的利润来抵消成本。

基于订阅的接入模式

✂ 许多服务运行商提供基于订阅的接入服务，用户每月或每年付固定的费用来换取接入服务。典型应用是数据库接入服务，包括论文、新闻、在线游戏等。但是很多用户并不想为浏览网上的内容而付费。

预付费接入模式

✂ 用户为一定时间段或一定内容的接入服务付一定金额的费用，还可以继续充值。预付费模式给用户在服务的花费方面以更大的控制权。

经纪人模式

✂ 经纪人是市场的创建者。作为中介，他们把买方和卖方凑在一起并鼓励他们交易，既可以是 B2C 和B2B 市场，也可以是 C2C 市场。经纪人通过其促成的每一笔交易或者按交易额的某百分比收取一定费用来赚钱。如很多团购网站。

广告商模式

✂ 广告商模式在电子商务中起着非常重要的作用。站点提供免费的接入但在每一个页面上都有广告，用户在一条广告商点一下就能进入广告商的页面，广告商按广告的展示方式和展示面积付费，或按广告的点击率付费。

门户网站模式

✂ 一个门户提供一站式的具体内容和服务的接入，比如新闻、股票信息、天气预报、聊天室等。还允许用户建立个性化的界面和内容，使访客更容易识别门户，门户网站可以更有针对性地投放广告。

免费接入模式

✂ 免费给用户一些东西，但是得看广告。如一些免费的 Web 空间提供商，它们在站点的顶端和底端提供了广告标题，有些电子贺卡带着个人信息和广告一起发送。

虚拟商场模式

✂ 虚拟商场是许多商家、服务提供商、经纪人和其他企业的托管站点，虚拟商场的运行商以建立和维

护商家的“摊位”和把商家编入站点目标而收取费用，也可以从商家成交的每一笔交易中收取费用。虚拟商场可以在更大型的站点内部运行。

虚拟社区模式

✂ 虚拟社区是一个吸引了网络上志同道合者一起组成的团体。用户分享信息，也以自己的方式贡献内容。由于用户自己都有所投入，所以他们对站点有很高的忠诚度，会定期访问，这为广告提供了可能性。如开心网、天涯社区等。

信息中介模式

✂ 信息中介收集、评价客户及其购物行为等信息，并把它们出售给那些想要接近客户的其他人。最初，给访客免费提供一些信息，而信息中介则对访客的在线活动进行监控。收集的信息对市场营销极有价值，它需要提供一些免费的信息吸引访客注册，收集大量对于针对性的广告活动有潜在价值的信息。

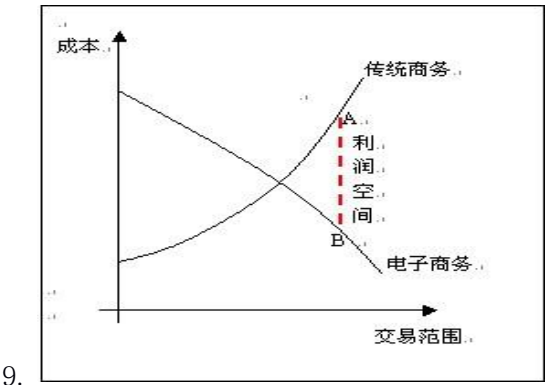
优缺点:

优势:

1. 突破了时空的限制
2. 降低成本
3. 同等竞争的机会，减少差异化
4. 更快更好的客户服务
5. 快速购物比较
6. 提升生产率和流通效率
7. 信息共享、便利和控制
8. 个性化定制

问题和束缚:

1. 前期投入
2. 安全性
3. 系统与数据的完整性和可扩展性
4. 客户关系的维护
5. 存在一些不适合在线购买的产品
6. 文化、语言和信任问题
7. 公司脆弱性（泡沫）和高风险
8. 法律法规不完善



域名及域名分类

域名：Domain Name，是由一串用点分隔的字符组成的在Internet上某一台计算机或计算机组的名称，用于在数据传输时标识计算机的电子方位。（物理位置）

域名就是因特网主机的地址，由它可转换为该主机在因特网中的物理位置（IP地址）

域名是主机所属单位在网络空间中的永久地址和名称

www.baidu.com

www.csdn.net

域名命名规则：

域名中包含以下字符：

26 个英文字母

0-9 十个数字

“-” 英文中的连字符

域名中字符的组合规则为：

不区分英文字母的大小写；

对于一个域名的长度有一定限制。（256个字符）

域名分类

从地域范围分类，可以分为[国际域名](#)和[国家域名](#)：

(1) 国际域名(International Top-Level Domain-names, ITLDs)：如表示工商企业的.com，表示网络提供商的.net，表示非盈利组织的.org等。

net	互联网络、接入网络的信息中心(NIC)和运行中心(NOC)
com	工、商、金融等企业
edu	教育机构
org	各种非盈利性的组织

mil	军事机构
ac	科研机构
info	提供信息服务的企业
gov(中国)	政府部门
...	...

(2) 国家域名(National Top-Level Domain-names, NTLDs): 按照国家不同分配不同后缀, 为国家的国内顶级域名。目前200多个国家和地区都按照ISO3166国家代码分配了顶级域名。

cn	中国
jp	日本
us	美国
de	德国
uk	英国
kr	韩国
tw	台湾
hk	香港
...	...

从域名的级别进行分类, 可分为[顶级域名](#)和[二级域名](#)。

(1) 顶级域名: 国家顶级域名、国际顶级域名。

(2) 二级域名: 二级域名是指在顶级域名下的域名。在国际顶级域名下, 它是指[域名注册人的网上名称](#), 如ibm(www.ibm.com)、microsoft、baidu、sohu等; 在国家顶级域名下, 它表示注册企业类别, 如edu.cn、gov.cn、com.cn等。

行政区域名有 34 个, 分别对应于我国各省、自治区和直辖市

虚拟主机与服务器

网站空间(Web Host), 又称为虚拟主机空间或虚拟主机, 指存放网站内容的计算机空间。

虚拟主机

把一台运行在互联网上的服务器划分为多个“虚拟”的服务器, 每一个虚拟主机都具有独立的域名和完整的Internet服务器功能。

利用虚拟主机技术, 一台服务器上的不同虚拟主机是各自独立的, 并由用户自行管理。

一台服务器主机只能支持一定数量的虚拟主机, 当超过这个数量时, 用户将会感到系统性能急剧下降。

虚拟主机的优势:

大幅降低费用 (购买服务器、租用专线、维护费用等)

性能相对较好 (合租服务器)

无需专门的维护人员和管理人员

服务器:

一个管理资源并为用户提供服务的计算机软件, 通常分为文件服务器(File Server)、数据库服务器(Database Server)和应用程序服务器(Application Server)。运行以上软件的计算机或计算机系统也称为服务器。

高速运算能力(cpu、内存)、海量的存储空间(硬盘)、长的可靠运行时间(主板)、强大的外部数据吞吐能力(网速)

服务器存储和处理网络上80%的数据, 被称为网络的灵魂。

网站服务器的性能、稳定性直接决定运行在其上的网站运行质量和运行效果。

服务器分类:

1. 塔式服务器

应用最为广泛, 最为常见的一种服务器。

类似普通台式机, 体积更大。

适用: 服务器扩展、散热性能要求较高, 采购数量不多, 且空间比较冗余。

优点

单个成本较低

由于机箱较大, 具备良好的扩展能力和散热能力

可以配置多路处理器、多根内存、多块硬盘、多个冗余电源和散热风扇

缺点

机器重量（最重）
空间占用率（最高）

机箱较大，配置可以很高，冗余扩展很齐备，应用非常广泛。
入门级服务器
能够满足大部分企业用户的要求

2. 机架式服务器

安装在机架上的服务器” 统一安装在按照国际标准设计的机柜中

优点
相对塔式服务器大大节省了空间占用， 使布线、管理更为简洁
节省了机房的托管费用
随着技术的发展，有着不逊色于塔式服务器的性能
平衡了性能和空间

缺点
由于机身的限制，在扩展能力和散热能力上不如塔式服务器
对制冷要求较高
需要对机架式服务器的体系结构专门进行设计，如主板、接口、散热系统等， 设计成本较高，价格一般也高于塔式服务器

适用于资金较为充裕，针对性比较强的应用，如需要密集型部署的服务运营商、群集计算等。
服务器第三方托管（如电信托管）的企业，按照空间(U)来收取托管费

3. 刀片服务器

在标准高度的机箱上插装多个卡式的服务器单元
服务器单元外观很薄，故名“刀片”，只有机架式服务器的1/3-1/2
每个刀片是一台独立的服务器，具有独立的CPU、内存、I/O总线，通过外置磁盘（板载硬盘）可以独立安装操作系统

优点
1. 扩展方便，支持热插拔
2. 通过刀片架组成服务器集群，提供高速的网络服务
3. 每个刀片不需要独立的电源、显示器、制冷装置，降低功耗，减少整体成本

缺点
没有形成统一的标准，几大巨头互不兼容，导致用户选择的范围很狭窄

网络营销概念及常见类型

网络营销(On-line Marketing或E-Marketing)就是以国际互联网络为基础，利用数字化的信息和网络媒体的交互性来辅助营销目标实现的一种新型的营销方式
网络营销就是以互联网为主要手段进行的、为达到一定营销目的的营销活动

类型

1.病毒营销

又称为病毒式营销，是指利用网民的主动在线传播，让某个带有广告信息的资讯像病毒一样传播出去，从而达到网络营销的目的

- 病毒营销特点
- 有吸引力的病源体
- 几何倍数的传播速度
- 高效率的接收
- 更新速度快

- 病毒营销传播方式（病毒载体）
- 图片
- 免费的电子贺卡、网页
- 免费下载的视频或Flash动画
- 免费下载的软件
- 免费下载的电子书
- 免费文章

论坛营销

企业利用论坛这一类型的网络交流平台，通过文字、图片、视频等方式发布企业的产品和服务的信息，从而让目标客户更加深刻地了解企业的产品和服务，最终达到企业宣传企业的品牌、加深市场认知度的网络营销活动

- 论坛营销优势
- 推广效果长久
- 营销费用低
- 受众基数大、影响面大
- 精准定位客户
- 互动性强、推广效果好
- 带来大量的外部链接

博客营销

- 企业博客营销形式
- 企业网站博客频道模式
- 建立在第三方博客企业博客平台的博客营销模式
- 个人独立博客网站模式
- 博客营销外包模式
- 博客广告模式

微博营销

微博营销以微博作为营销平台，每一个听众（粉丝）都是潜在的营销对象，企业利用更新自己的微博向网友传播企业和产品信息，树立良好的企业形象和产品形象，从而达到营销的目的。

微信营销

微信营销是网络经济时代企业营销模式的一种，是伴随着微信的火热而兴起的一种网络营销方式
微信不存在距离的限制，用户注册微信后，可与周围同样注册的“朋友”形成一种联系，订阅自己所需的信息，商家通过提供用户需要的信息，推广自己的产品，从而实现点对点的营销

软文营销

所谓软文，就是指通过特定的概念诉求、以**摆事实讲道理的方式使消费者走进企业设定的"思维圈"**，**以强有力的针对性心理攻击迅速实现产品销售的文字模式。**

狭义的：指企业门户自主花钱在报纸或杂志等宣传载体上刊登的纯文字性的广告。

广义的：指企业/门户通过策划在报纸、杂志、DM（直邮）、网络、手机短信等宣传载体上刊登的可以提升企业品牌形象和知名度，或可以促进企业/网站的产品销售的一些宣传性、阐释性文章等。

邮件营销

定义：

邮件营销(E-mail 营销)是指在**用户实现允许的前提条件下**，通过电子邮件方式向目标用户**传递有价值的信息**进行网络营销的一种手段，也称为“许可邮件营销”
要注意与垃圾邮件的区别

邮件营销的效果评价：

退信率（Bounce rate）：没有送达的邮件的比率，该比率是评价列表质量的一个重要指标。
开信（浏览）率（Open rate，View rate）：用户在收到信件后打开阅览的比例。
点击率（Click-through rate，CTR）：用户收到信件后点击其中的链接进入广告主指定网页的比例。许可电子邮件的 CTR 可以达到10%左右，而租用名单发送的电子邮件的 CTR 通常在 1%到 2%之间。
转化率（Conversion rate）：用户收到信件后产生购买、用户注册、

期刊订阅等预期行动的比例。

新顾客获得率 (Acquisition rate)：收到信件的用户转化为公司新顾客的比率。该比率用来评价列表的质量和促销的效果。

退订率 (Unsubscribe rate)：用户收到了邮件但是要求退订的比率。该比率可以用来评价营销信息的质量以及发送的频率是否恰当。正常的退订率在 1%以下。

投资回报率 (ROI)；每次销售成本 (Cost per sale)；每次回应成本 (Cost per response)；每条讯息成本 (Cost per message) ……

4Cs营销理论

4Cs理论是由美国营销专家罗伯特·劳特朋教授针对4Ps营销理论存在的问题于1990年提出来的，即顾客(Customer)、成本(Cost)、便利(Convenience)、沟通 (Communication) 组成的营销手段

4Cs营销理论以消费者需求为导向，重新设定了市场营销组合的四个基本要素：

瞄准消费者的需求和期望 (Customer)、
消费者所愿意支付的成本 (Cost)、
消费者购买的方便性 (Convenience)、
与消费者沟通 (Communication)

网络营销的模式是从消费的需求出发，营销决策是在满足 4Cs

的要求的前提下的企业利润最大化，最终实现的是消费者需求的

满足和企业利润最大化

搜索引擎定义

搜索引擎(Search Engine)是指根据一定的策略、运用特定的计算机程序收集互联网上的信息,在对信息进行组织和处理后,并将处理后的信息显示给用户,是为用户提供检索服务的系统。

搜索引擎分类

全文搜索引擎

目录索引

目录索引，顾名思义就是将网站分门别类地存放在相应的目录中，因此用户在查询信息时，可选择关键词搜索，也可按分类目录逐层查找。

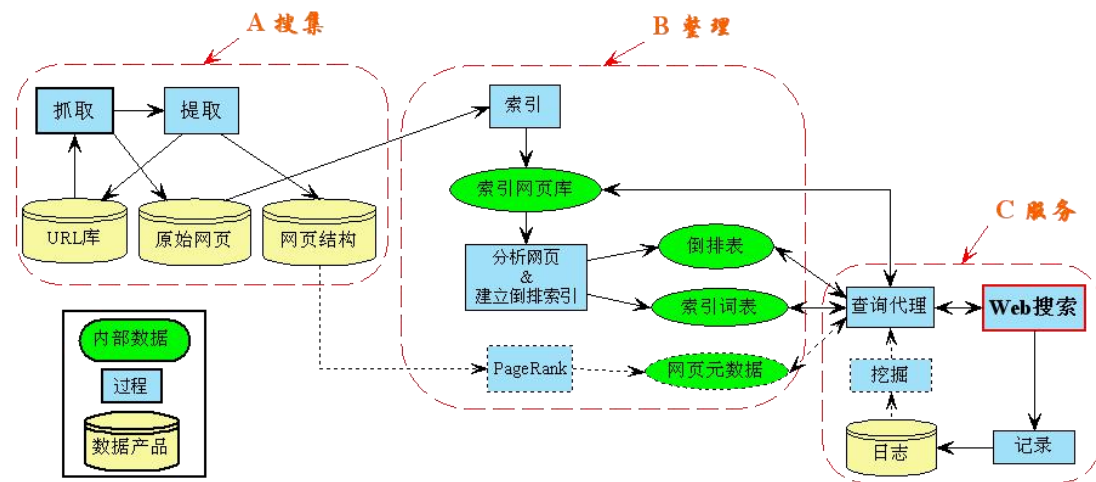
元搜索引擎

元搜索引擎（Meta Search Engine）接受用户查询请求后，同时在多个搜索引擎上搜索，并将结果返回给用户。

垂直搜索引擎

垂直搜索专注于特定的搜索领域和搜索需求（例如：机票搜索、旅游搜索、生活搜索、小说搜索、视频搜索等等），
在其特定的搜索领域有更好的用户体验。相比通用搜索动辄数千台检索服务器，垂直搜索需要的硬件成本低、用户需求特定、查询的方式多样。

搜索引擎工作原理



1. 抓取网页

每个独立的搜索引擎都有自己的**网页抓取程序 (spider, 蜘蛛程序)**。Spider顺着网页中的超链接,连续地抓取网页,被抓取的网页被称之为**网页快照**。由于互联网中超链接的应用很普遍,理论上,从一定范围的网页出发,就能搜集到绝大多数的网页。

2. 处理网页

搜索引擎抓到网页后,还要做大量的预处理工作,才能提供检索服务。其中,最重要的就是**提取关键词,建立索引文件**,其他还包括**去除重复网页、分词 (中文)、判断网页类型、分析超链接、计算网页的重要度/丰富度**等。

3. 提供检索服务

用户输入关键词进行检索,搜索引擎**从索引数据库中找到匹配该关键词的网页**;为了用户便于判断,除了网页标题和URL外,还会提供一段来自网页的摘要以及其他信息。

搜索引擎组成部分

搜索引擎一般由**搜索器、索引器、检索器和用户接口**四个部分组成:

搜索器: 其功能是在互联网中漫游,发现和搜集信息;

索引器: 其功能是理解搜索器所搜索到的信息,从中抽取出索引项,用于表示文档以及生成文档库的索引表;

索引表一般使用某种形式的倒排表 (Inversion List), 即由索引项查找相应的文档。索引表也可能要记录索引项在文档中出现的位置,以便检索器计算索引项之间的相邻或接近关系 (proximity)。

索引器可以使用**集中式索引算法或分布式索引算法**。当数据量很大时,必须实现即时索引 (Instant Indexing), 否则不能够跟上信息量急剧增加的速度。索引算法对索引器的性能 (如

大规模峰值查询时的响应速度) 有很大的影响。一个搜索引擎的有效性在很大程度上取决于索引的质量。

检索器: 其功能是根据用户的查询在索引库中快速检索文档, 进行相关度评价, 对将要输出的结果排序, 并能按用户的查询需求合理反馈信息;

用户接口(HTML页面): 其作用是接纳用户查询、显示查询结果、提供个性化查询项。

中文分词算法

方法:

1. 基于词典的分词方法

与词典、词条进行匹配.

最大匹配法: 正向最大匹配, 逆向最大匹配, **双向最大匹配**(BMM)

2. 基于统计的分词方法

如果**相连的字在不同的文本中出现的次数越多**, 就证明这相连的字很可能就是一个词

基于统计的分词方法是在给定大量已经分词的文本的前提下, 利用统计机器学习模型学习词语切分的规律 (称为训练), 从而实现对未知文本的切分。

3. 基于理解的分词方法

基本思想是在分词的同时进行语法、语义分析, 利用语法信息和语义信息来处理歧义现象

它通常包括三个部分: 分词子系统、句法语义子系统、总控部分。在总控部分的协调下, 分词子系统可以获得有关词、句子等的句法和语义信息来对分词歧义进行判断, 即它模拟了人对句子的理解过程。

难点:

交叉型歧义、组合型歧义、混合型歧义(包含了前两种.)

新词识别很难: 一是指已有的词表中没有收录的词; 二是指已有的训练语料中未曾出现过的词

jieba (结巴分词): 国内使用人数最多的中文分词工具

SnowNLP: SnowNLP是一个python写的类库, 可以方便的处理中文文本内容

THULAC: 由清华大学自然语言处理与社会人文计算实验室研制推出的一套中文词法分析工具包

NLPIR: NLPIR分词系统 (前身为2000年发布的ICTCLAS词法分析系统), 是由北京理工大学张华平博士研发的中文分词系统, 经过十余年的不断完善, 拥有丰富的功能和强大的性能

所谓锚文字, 就是页面内某个出链附近的一些描述文字

链接分析方法(pagerank算法)

基本思想:

认可度越高的网页越重要, 即反向链接(back link)越多的网页越重要

反向链接的源网页质量越高, 被这些高质量网页的链接指向的网页越重要

链接数越少的网页越重要

计算过程:

假设一个由4个网页组成的群体: A, B, C和D。如果所有页面都只链接至A, 那么A的PR(PageRank)值将是B、C及D的PageRank的总和。

$$PR(A) = PR(B) + PR(C) + PR(D)$$

重新假设B链接到A和C, C只链接到A, 并且D链接到全部其他的3个页面, 一个页面总共只有一票。所以B给A和C每个页面半票。以同样的逻辑, D投出的票只有三分之一算到了A的PageRank上。

$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3}$$

PageRank公式:

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

n: 网络中网页总数

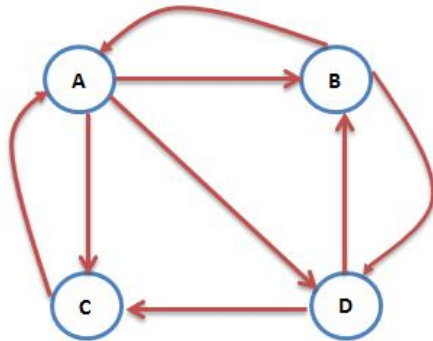
d: 阻尼因子, 即按照超链接进行浏览的概率

1-d: 随机跳转一个新网页的概率

PR(Pj):网页Pj的PR值

L(Pj): 网页Pj的链出网页数

pagerank例:



这个例子中只有四个网页，如果当前在A网页，那么悠闲的上网者将会各以1/3的概率跳转到B、C、D，这里的3表示A有3条出链，如果一个网页有k条出链，那么跳转任意一个出链上的概率是1/k，同理D到B、C的概率各为1/2，而B到C的概率为0。一般用转移矩阵表示上网者的跳转概率，如果用n表示网页的数目，则转移矩阵M是一个n*n的方阵；如果网页j有k个出链，那么对每一个出链指向的网页i，有 $M[i][j]=1/k$ ，而其他网页的 $M[i][j]=0$ ；上面示例图对应的转移矩阵如下：

$$M = \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

(从上指到左)

初试时，假设上网者在每一个网页的概率都是相等的，即1/n，于是初试的概率分布就是一个所有值都为1/n的n维列向量 V_0 ，用 V_0 去右乘转移矩阵M，就得到了第一步之后上网者的概率分布向量 MV_0 ， $(n \times n) * (n \times 1)$ 依然得到一个 $n \times 1$ 的矩阵。下面是 V_1 的计算过程：

$$V_1 = MV_0 = \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 9/24 \\ 5/24 \\ 5/24 \\ 5/24 \end{bmatrix}$$

不断左乘 转移矩阵

注意矩阵M中M[i][j]不为0表示用一个链接从j指向i，M的第一行乘以V0，表示累加所有网页到网页A的概率即得到9/24。得到了V1后，再用V1去右乘M得到V2，一直下去，最终V会收敛，即 $V_n = MV_{(n-1)}$ ，上面的图示例，不断的迭代，最终 $V = [3/9, 2/9, 2/9, 2/9]$ ：

$$\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \begin{bmatrix} 9/24 \\ 5/24 \\ 5/24 \\ 5/24 \end{bmatrix} \begin{bmatrix} 15/48 \\ 11/48 \\ 11/48 \\ 11/48 \end{bmatrix} \begin{bmatrix} 11/32 \\ 7/32 \\ 7/32 \\ 7/32 \end{bmatrix} \dots \begin{bmatrix} 3/9 \\ 2/9 \\ 2/9 \\ 2/9 \end{bmatrix}$$

SEO相关术语

SEO (Search Engine Optimization), 译为搜索引擎优化, 就是针对搜索引擎对网页的检索特点, 让网站建设的各项基本要素适合搜索引擎的检索原则, 从而使搜索引擎收录本网站尽可能多的网页, 并且在搜索引擎自然检索结果中排名靠前, 最终达到网站推广的目的。

PV (Page View):

PV是用户访问网站时所浏览的总网页数。平均PV值实际上从侧面反映出了一个网站的用户粘性，即网站受欢迎程度。

百度快照:

百度快照其实就是百度蜘蛛在爬取网页内容时，将网页文本内容独立创建一个文本文件，保存在百度数据库中的网页的“照片”。

SEO黑帽 (SEO black hat) :

简单来说，SEO黑帽就是采用了搜索引擎所禁止的方法（作弊），对网站在搜索引擎的排名进行优化，在短时间内排到搜索引擎前列。

白帽所指的是采用合理的SEO行为优化网站，提高用户的体验，合理与其他网站互联，从而达到网站在搜索引擎排名的稳步提升。

跳出率

是指来到网站只浏览了一个页面就离开的用户占网站总浏览用户的百分比。

沙盒效应(Sandbox):

指一个新建的网站在刚上线的一个或几个月内，在搜索引擎上很难有好的排名，甚至没有排名。

SPAM(Specially Processed Assorted Meat):

在搜索引擎中的SPAM是指专门针对那些欺骗搜索引擎的技术。

Site:网址 查找被搜索引擎收录的网页数目

常用SEO技术

网页title的优化(★★★★★)

title 标签具有不可替代性，是网页第一个重要标签，也是搜索引擎了解网页的入口，它是对网页主题归属的最佳判断点。

关键词由小到大，由细到粗，将整个网页的内容都毫无保留地描述在title中，达到便于搜索引擎蜘蛛索引的目的。

考虑到整个网页的关键词密度，页面的关键词一般只在title中出现一次即可，否则会有关键词堆砌的作弊嫌疑。

如果要单独优化百度或者其他国内搜索引擎，就使用下划线(“_”)来分割title关键词；

如果优化Google或者其他国外的搜索引擎，就采取用横杆(“-”)作为分割符。

如果想要在Google和百度都有良好表现的话，建议还是使用下划线(“_”)。

不建议使用“|”作为分割符。

关键词的选取(★★★★★)

找到一个合适自己网站的关键词，是整个SEO过程中最为关键的一步。

关键词不要太宽泛

长尾关键词并不是很长的关键词（意义要精确）

多考虑用户会以什么关键词搜索

选择竞争较小，但实际关注度高的关键词

关键词和网站一定要有关系

关键词的优化(★★★★★)

元标签关键词的优化

元标签就是HTML中<head>与</head>中间的<meta>标签。

其作用是告诉一切支持网页运行的程序或者搜索引擎蜘蛛该网页的内容和基本信息。

meta标签包含很多属性，例如关键词、描述、版权等。

对于搜索引擎而言，蜘蛛爬取时网页首先要了解这个网页的内容是什么，属于什么类别，因此，在网页上设置合理的meta标签非常重要。

meta标签最常用的方式是：<meta name= “keywords”>

标准的写法是在每个关键词之间以英文的逗号(,)分隔

在确定关键词后必须开始控制网页的关键词密度，即出现的频率。如果关键词密度过低，那么对于优化起不到重要的作用；而如果密度过高，又会被搜索引擎认为是作弊。

元标签及网页描述的优化(★★★★☆)

最常用的元标签包括keywords和description。

如果一个网页的description写法不规范或者没有description，搜索引擎索引并显示的内容是整个网页前面一些字符；如果description非常标准的话则显示结果就是其中的内容。

description在网页排序中有很重要的作用。

description可以使用长达255个字符来描述一个网页。

正确写法：

最好用一句话来概括本网页的主题内容。

不要超过255个字符，搜索引擎只会索引前255个字符。

使用英文格式的标点符号。

robots.txt文件(★★★★☆)

robots.txt必须放置在一个站点的根目录下，而且文件名必须全部小写。

All: 不拒绝任何蜘蛛的抓取行为

Noindex: 不索引当前页面

Index: 索引当前页面

Nofollow: 不追踪当前网页的链接

Follow: 追踪当前网页的链接

Noarchive: 在搜索结果中不保存当前网页的快照

Archive: 在搜索结果中保存当前网页的快照

Nosnippet: 在搜索结果中不采用当前页面的头部描述信息，且不保存当前页面的快照。

None: 不索引当前页面以及当前页面的所有链接。

如果包含多个关键词，需要在中间以英文的逗号分隔。

不可以两个反义词都用。

Googlebot	Google对一般网页的搜索蜘蛛
Googlebot-Mobile	Google对于移动设备，例如手机网页的索引蜘蛛
Googlebot-Image	Google专门用来抓取图片的蜘蛛
Baiduspider	百度综合索引蜘蛛
Yahoo!Slurp	雅虎综合索引蜘蛛
YodaoBot	网易有道综合索引蜘蛛
Sosospider	腾讯搜搜综合索引蜘蛛
Sogou spider	搜狐搜狗综合索引蜘蛛

```
User-agent: *  
Disallow:  
或者  
User-agent: *  
Allow: /
```

允许所有搜索引擎访问所有内容

✓ robots.txt参考写法5

```
User-agent: *  
Disallow: /css/  
Disallow: /admin/  
Disallow: /images/
```

禁止搜索引擎索引指定目录内容

♦ robots.txt文件(★★★★☆)

✓ robots.txt参考写法8

```
User-agent: *  
Allow: .asp$  
Disallow: /
```

通配符\$的使用，允许访问所有以asp为后缀的文件

✓ robots.txt参考写法9

```
User-agent: *  
Disallow: /*? *
```

禁止访问所有URL中包含?的网页，如index.jsp?id=1

网站结构和URL优化(★★★★☆)

如果URL过长，或者带有“?”一类的参数，搜索引擎会认为该网站不友好。
URL结构中，三层目录是最为合适的。

Google网站管理员帮助：简化域名后面的参数，如果可以采取伪静态或者直接生成静态网页的方式，会对搜索引擎索引有很大的帮助。

选择一个适合网站内容的域名

尽量选择包含关键词的域名，如chinauml.com或chinauml.net，必要的时候在中间加横杆，如china-uml.com或china-uml.net。

在注册一个域名之前，了解该域名是否曾经被人使用过，有没有被搜索引擎惩罚过。

二级域名和主域名的权重是一样的，将分散SEO精力，可能会让每个二级域名下的网页排名都不会很好。

网页链接的优化(★★★★☆)

内部链接的优化

网站导航的优化

网站导航尽量使用文本形式来作为链接的锚文字。

锚文字——“组合模式”

尽量不使用JavaScript脚本、图片和Flash动画等作为网站导航。

如果一定要使用，注意如下几点：

- (1) 在页面的其他部位(如导航条下部、页面底部等)放置文本文字的链接，链接内容与Flash、图片和JS导航内容相同。
- (2) 在页面的底部或者其他显著位置放置HTML格式（或XML格式）的网站地图sitemap。该sitemap包含各个频道、栏目的链接。
- (3) 图片导航添加alt属性。

小技巧：

- (1) 可以通过查询域名whois信息来确定网站的年龄，通过“site:域名”查看网站收录网页的数量。
- (2) 查询网站的PR值，如<http://pr.chinaz.com/>。
- (3) 网站对外的链接越少，对网站自身的好处越大；链向网站的链接越多，对网站自身的好处越大。

Heading标签的优化(★★★★☆)

根据Google官方SEO手册解释，heading标签就是指六种不同大小的文字，从<h1>到<h6>，文字由大到小。

在SEO中,heading标签有着举足轻重的地位。

蜘蛛在爬行过程中，首先会通过链接向该网页的链接**锚文字**获取“**第一印象**”，以大致判断该网页的内容。

然后**查询网页title标签**的内容，以确认刚才**查看锚文本时的判断是否正确**。

第三步就是网页的h1标签，h1标签作为**整个网页的内容标题**，可以概括整个页面的内容。蜘蛛会以h1的内容作为参考对网页本身进行分类。

从h1到h6，搜索引擎的关注度逐渐降低。

图片优化(★★★★☆)

目前主流的图片搜索引擎中，大多数都是以文本链接或图像的alt属性等HTML元素来辨别图像内容，当然也有根据图片本身的内容来做图像内容发掘。

通过文本等第三方因素来辨别图片内容的图像搜索引擎(大部分)，根据链接的锚文本、title标签、heading标签、alt属性等来判断图片的内容。

通过分析图片本身来确定图形内容的搜索引擎（不多），例如华盛顿大学的retrievr。

(1) 通过HTML代码中的元素提高图片与关键词之间的相关度。

(2) 通过URL链接来提高图片与关键词的相关度

(3) 通过文件名、URL文件夹名称进行图片SEO

(4) 环绕图片的文字对搜索引擎蜘蛛的干扰。

网页中的图片标签的前255个字符和后255个字符对搜索引擎图片内容的判断有着非常重要的作用。

与普通的网页排序一样，图片搜索引擎的排序和图片与关键词的相关度也有很重要的联系。用户搜索的关键词与网站上的图片的相关度越高，排序也就越靠前。

通过合理的手段将图片所属网页、图片本身的内容与关键词之间的相关度提高，那么网站的图片在百度等图片搜索引擎中也会取得更好的排名。

网页减肥(★★★☆☆)

如何让蜘蛛能够轻盈地抓取网页的核心内容？

JavaScript脚本代码的减肥：使用js文件封装JavaScript代码，调换HTML代码中JavaScript的位置，尽量将JS代码放在HTML代码之后。

网页文本的减肥：分页处理

网页图片的减肥：图片的剪裁和压缩

Flash的减肥：尽量避免使用Flash，制作与Flash中内容相同的文字网页。

SEO常用作弊方式

关键词堆砌

桥页

隐藏文字是在网页的HTML文件中放上含有关键词的文字，但这些文字不能被用户所看到，只能被搜索引擎看到。

隐藏链接(Hidden Link)

隐藏链接和隐藏文字相似，但是区别是把关键词放在链接里面，而这个链接也是用户所看不到的。

隐藏页面(Cloaked Page)

有的网页使用程序或脚本来检测来访问的是搜索引擎还是普通用户。

数据挖掘及核心任务

定义:

从大量数据中寻找其规律的技术, 是统计学、数据库和人工智能等技术的综合, 是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中, 提取隐含在其中的、人们事先不知道的、但又是潜在的有用信息和知识的过程

数据挖掘的四大核心任务

关联分析 (Association Analysis)

某种事物发生时其他事物也会发生, 这种联系称之为关联, 是指两个或多个变量取值之间存在的一类重要的可被发现的某种规律性

用来发现数据中强关联特征的模式

关联规则挖掘方法

设定最小支持度和置信度的阈值: minsup和minconf

目标: support \geq minsup 且 confidence \geq minconf

找出所有的频繁项集

由频繁项集产生强关联规则

例如:

总项集个数为 2^d

$$R = \sum_{k=1}^{d-1} \left[\binom{d}{k} \times \sum_{j=1}^{d-k} \binom{d-k}{j} \right]$$
$$= 3^d - 2^{d+1} + 1$$

关联规则的总个数为

分类与预测 (Classification & Prediction)

用于预测离散的目标变量，通过分析一个类别已知的数据集的特征来建立一组模型，该模型可用以**预测类别未知的数据项的类别**(分类预测的是类别，而非连续的数值)

预测:

用于预测连续的目标变量，预测也是首先构建模型，再通过模型来预测未知值，主要方法是回归(Regression)，包括线性回归、多元回归和非线性回归等

聚类分析 (Clustering Analysis)

无监督的分类，是指把一组数据分成不同的“簇”，每簇中的数据相似而不同簇间的数据则距离较远

保证不同类间数据的相似性尽可能地小，而类内数据的相似性尽可能地大

异常检测 (Anomaly Detection)

偏差分析 (Deviation Analysis) 或离群点分析 (Outlier Analysis)

离群点 (outlier): 异常对象，属性值明显偏离期望的或常见的属性值

应用于欺诈检测、入侵检测、医疗检查等

常用方法

基于模型的技术 (概率分布模型)

基于邻近度的技术 (基于距离的离群点检测技术)

基于聚类技术

基于密度的技术

常用数据挖掘算法:

Apriori算法

关联分析算法

几个概念:

项集: 一个或多个项的集合

支持数: 一个项目所在交易中出现的次数

支持度: 支持数/总交易数(与方向无关)

置信度: $c(X \rightarrow Y)$ 同时包含x和y的交易数/包含x的交易数.

设定最小支持度和置信度的阈值: minsup 和 minconf

目标: $\text{support} \geq \text{minsup}$ 且 $\text{confidence} \geq \text{minconf}$

找出所有的频繁项集

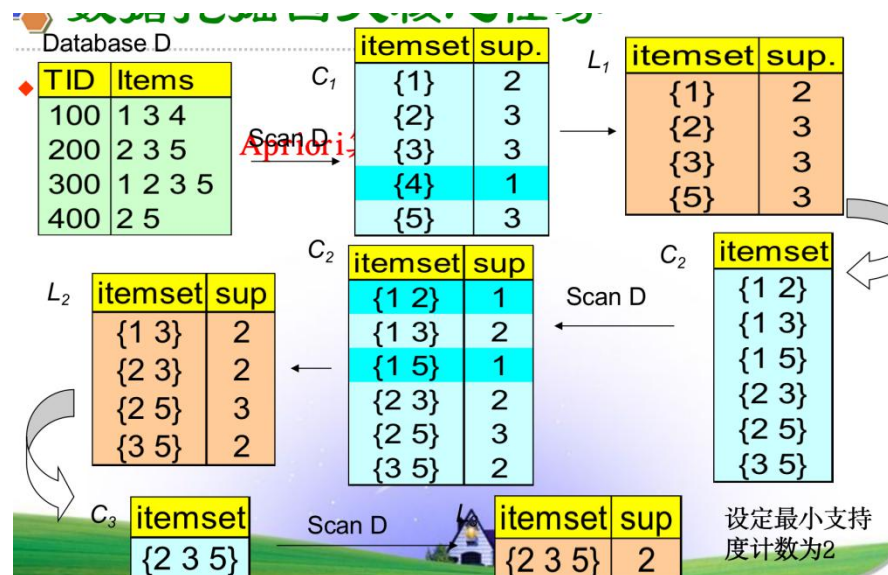
由频繁项集产生强关联规则

基本思想:

频繁项集的子集也一定是频繁的, 如果 $\{A, B\}$ 是频繁项集, 则 $\{A\}$ 和 $\{B\}$ 也一定是频繁项集

从1到k (k项集) 递归查找频繁项集

用得到的**频繁项集生成关联规则**



算法缺陷

产生大量的候选集

可能需要重复扫描数据库

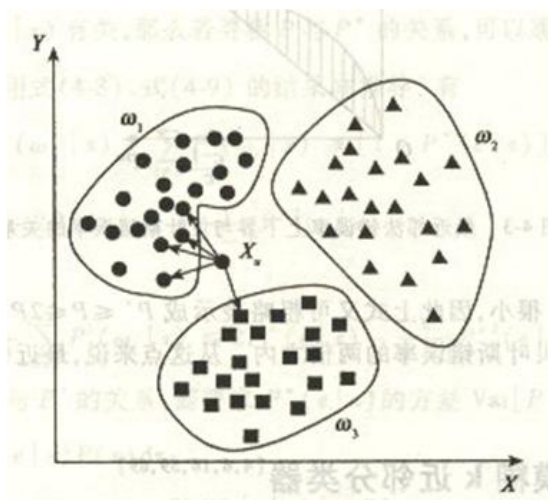
Knn算法

K-Nearest Neighbor

所谓 K 最近邻，就是 k 个最近的邻居的意思，说的是每个样本都可以用它最接近的 k 个邻居来代表。

分类算法

找k个最近的邻居,这个k个里面,可能属于不同的类别,按照少数服从多数,哪个类别出现的次数多就属于哪个类别。



优点

简单，易于理解，易于实现，无需估计参数，无需训练；

适合对稀有事件进行分类；

特别适合于多分类问题(multi-modal,对象具有多个类别标签)， kNN 比 SVM 的表现要好。

缺点

该算法在分类时有个主要的不足是，当样本不平衡时，如一个类的样本容量很大，而其他类样本容量很小时，有可能导致当输入一个新样本时，该样本的 K 个邻居中大容量类的样本占多数。该算法只计算“最近的”邻居样本，某一类的样本数量很大，那么或者这类样本并不接近目标样本，或者这类样本很靠近目标样本。无论如何，数量并不能影响运行结果。

该方法的另一个不足之处是**计算量较大**，因为对每一个待分类的文本都要计算它到全体已知样本的距离，才能求得它的 K 个最近邻点。

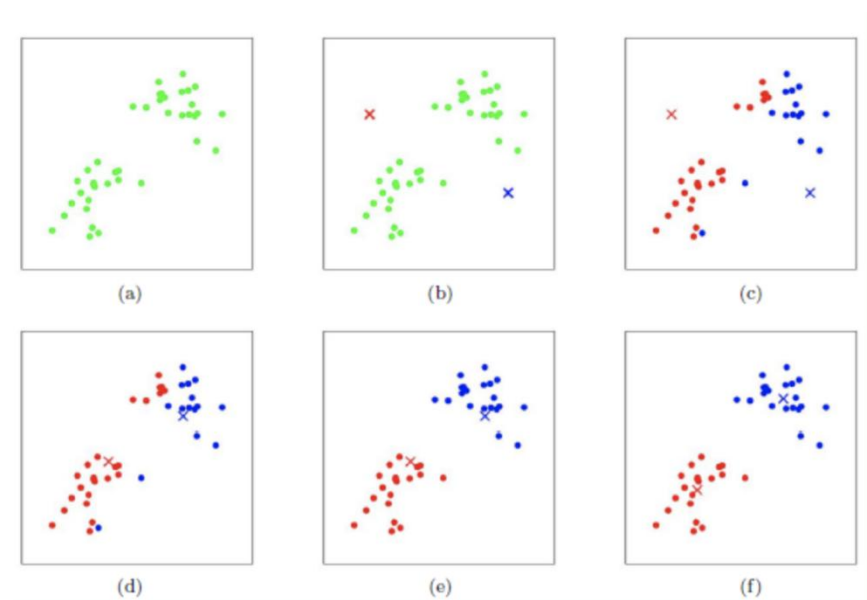
可理解性差，无法给出像决策树那样的规则。

K-means算法

聚类算法

首先，随机选择 K 个点做为聚类中心

然后，计算其它点到这些聚类中心点的距离，通过对簇中距离平均值的计算，不断改变这些聚类中心的位置，直到这些聚类中心不再变化为止



在图b中，我们随机选择了两个 k 类所对应的类别质心，即图中的红色质心和蓝色质心，然后分别求样本中所有点到这两个质心的距离，并标记每个样本的类别为和该样本距离最小的质心的类别，如图c所示，经过计算样本和红色质心和蓝色质心的距离，我们得到了所有样本点的第一轮迭代后的类别。此时我们对当前标记为红色和蓝色的点分别求其新的质心，如图4所示，新的红色质心和蓝色质心的位置已经发生了变动。图e和图f重复了我们在图c和图d的过程，即将所有点的类别标记为距离最近的质心的类别并求新的质心。最终我们得到的两个类别如图f。我们一般会多次运行图c和图d，才能达到最终的比较优的类别

步骤:

1. 从 n 个数据对象任意选择 k 个对象作为初始聚类中心;
2. 根据每个聚类对象的均值 (中心对象), 计算每个对象与这些中心对象的距离; 并根据最小距离重新对相应对象进行划分;
3. 重新计算每个 (有变化) 聚类的均值 (中心对象);
4. 计算标准测度函数, 当满足一定条件, 如函数收敛时, 则算法终止; 如果条件不满足则回到步骤 (2) 。

算法优点

算法快速、简单;

对大数据集有较高的效率并且是可伸缩性的;

时间复杂度近于线性, 而且适合挖掘大规模数据集。K-Means 聚类算法的时间复杂度是 $O(nkt)$, 其中 n 代表数据集中对象的数量, t 代表着算法迭代的次数, k 代表着簇的数目。

算法缺陷

① 在 K-means 算法中 K 是事先给定的, 这个 K 值的选定是非常难以估计的。很多时候, 事先并不知道给定的数据集应该分成多少个类别才最合适。这也是 K-means 算法的一个不足。(尽量 k 越小越好, 平衡距离, 找拐点。

2、聚类初始中心不好找, 这个初始聚类中心的选择对聚类结果有较大的影响, 一旦初始值选择的不好, 可能无法得到有效的聚类结果

③ 从 K-means 算法框架可以看出, 该算法需要不断地进行样本分类调整, 不断地计算调整后的新的聚类中心, 因此当数据量非常大时, 算法的时间开销是非常大的。

考试题型:

选择题 15个30分

三分一个? 哦不对 是两分一个.

判断题 10个10分

对错题..., 不用背感觉是

简答题 5个20分

这个就是要背的部分了...

押题小王子:

4Cs营销理论

三、名词解释（共 5 题，每题 4 分，共 20 分）

1. 电子商务
2. 域名
3. 搜索引擎
4. 数据挖掘
5. SEO

三、名词解释（每题 4 分，共 20 分）

1. 电子商务：电子商务(Electronic Commerce, 简称 EC)是指在全球各地广泛的商业贸易活动中，在因特网开放的网络环境下，基于浏览器/服务器应用方式，买卖双方不谋面地进行各种商贸活动，实现消费者的网上购物、商户之间的网上交易和在线电子支付以及各种商务活动、交易活动、金融活动和相关的综合服务活动的一种新型的商业运营模式。简言之，电子是手段，商务是目的。
2. 域名：域名(Domain Name)是由一串用点分隔的字符组成的在 Internet 上某一台计算机或计算机组的名称，用于在数据传输时标识计算机的电子方位。从因特网的管理角度看，域名就是因特网主机的地址，它可以转换为该主机在因特网中的物理位置（IP 地址）。
3. 搜索引擎：搜索引擎(Search Engine)是指根据一定的策略、运用特定的计算机程序收集互联网上的信息，在对信息进行组织和处理后，并将处理后的信息显示给用户，是为用户提供检索服务的系统。
4. 数据挖掘：数据挖掘(Data Mining)是从大量数据中寻找其规律的技术，是统计学、数据库和人工智能等技术的综合，是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在的有用信息和知识的过程。
5. SEO：SEO(Search Engine Optimization)，译为搜索引擎优化，是针对搜索引擎对网页的检索特点，让网站建设的各项基本要素适合搜索引擎的检索原则，从而使搜索引擎收录本网站尽可能多的网页，并且在搜索引擎自然检索结果中排名靠前，最终达到网站推广的目的。

综合题 4个40分

四个算法,每个十分.

通过用户口碑宣传网络，信息像病毒一样传播和扩散，利用快速复制的方式传向数以千计、数以百万计的受众。

(1) 病毒营销：是指利用网民的主动在线传播，让某个带有广告信息的资讯像病毒一样传播出去，从而达到网络营销的目的。病毒营销并非真的以传播病毒的方式开展营销，而是

(2) 论坛营销：企业利用论坛这一类型的网络交流平台，通过文字、图片、视频等方式发布企业的产品和服务的信息，从而让目标客户更加深刻地了解企业的产品和服务，最终达到企业宣传企业的品牌、加深市场认知度的网络营销活动。

(3) 博客营销：利用博客进行营销，它是企业利用博客这种特殊的交互性应用方式，发布企业信息，及时与客户沟通交流，开展市场活动的新型营销方式。

(4) 微博营销：以微博作为营销平台，每一个听众（粉丝）都是潜在的营销对象，企业利用更新自己的微博向网友传播企业和产品信息，树立良好的企业形象和产品形象，从而达到营销的目的。

(5) 微信营销：微信营销是网络经济时代企业营销模式的一种，是伴随着微信的火热而兴起的一种网络营销方式。微信营销主要体现在以安卓系统、苹果系统、Windows Phone 系统的手机或者平板电脑中的移动客户端进行的区域定位营销。

(6) 软文营销：指通过特定的概念诉求、以摆事实讲道理的方式使消费者走进企业设定的“思维圈”，以强有力的针对性心理攻击迅速实现产品销售的文字营销模式，包括悬念式、故事式、恐吓式、情感式和促销式等形式。

(7) 邮件营销：在用户实现允许的前提下，通过电子邮件方式向目标用户传递有价值的信息进行网络营销的一种手段，也称为“许可邮件营销”。

评分标准：正确列举六种网络营销手段计 6 分，简要说明正确计 4 分。

- 搜索引擎工作原理：

- (1) 抓取网页

- 每个独立的搜索引擎都有自己的网页抓取程序（spider，蜘蛛程序）。Spider 顺着网页中的超链接，连续地抓取网页，被抓取的网页被称之为网页快照。由于互联网中超链接的应用很普遍，理论上，从一定范围的网页出发，就能搜集到绝大多数的网页。

- (2) 处理网页

- 搜索引擎抓到网页后，还要做大量的预处理工作，才能提供检索服务。其中，最重要的就是提取关键词，建立索引文件。其他还包括去除重复网页、分词（中文）、判断网页类型、分析超链接、计算网页的重要度/丰富度等。

- (3) 提供检索服务

- 用户输入关键词进行检索，搜索引擎从索引数据库中找到匹配该关键词的网页；为了用户便于判断，除了网页标题和 URL 外，还会提供一段来自网页的摘要以及其他信息。

- 搜索引擎一般由搜索器、索引器、检索器和用户接口四个部分组成：

- (1) 搜索器：其功能是在互联网中漫游，发现和搜集信息；

- (2) 索引器：其功能是理解搜索器所搜索到的信息，从中抽取出索引项，用于表示文档以及生成文档库的索引表；

- (3) 检索器：其功能是根据用户的查询在索引库中快速检索文档，进行相关度评价，对将要输出的结果排序，并能按用户的查询需求合理反馈信息；

- (4) 用户接口(HTML 页面)：其功能是接纳用户查询、显示查询结果、提供个性化查询项。

通过这些 SEO 技术来优化网页以及在搜索引擎中提高排名。

(1) 关键词选取与密度控制：通过百度关键词指数工具和百度竞价排名关键词推荐工具等工具选取关键词；关键词不要太宽泛；选择长尾关键词；多考虑用户会以什么关键词搜索；选择竞争较小，但实际关注度高的关键词；关键词和网站一定要有关系；关键词密度控制在 2%-8% 之间。

(2) Title 标签：尽量简短，符合网页内容，不重复，不能堆砌关键词；关键词由小到大、由粗到细；注意关键词密度；使用下划线 (“_”) 作为分割符；不能超过 255 个字符。

(3) Meta 标签的 keywords 属性与 description 属性：keywords 中不能有太多关键词，使用英文逗号 (“,”) 作为分割符；description 中通过一句话来概括本网页的主题内容，不要超过 255 个字符，使用英文格式的标点符号。

(4) Heading 标签：从 h1 到 h6 标签搜索引擎的关注度由高到低；每个页面不能有超过 1 个 h1 标签，h2 标签不能太多，不要超过 3 个。

(5) URL 优化：尽量使用三层目录，选择一个适合网站内容的域名，URL 中最好带有关键词，实施页面静态化，减少使用自动生成的 URL，只用一个 URL 访问网页，尽量使用小写字母。

(6) 伪静态化：通过正则表达式的方式来对 URL 字符串进行转换，防止出现带有 “?” 一类的参数的 URL。

(7) 内部链接优化：尽量使用文字链接；尽量不使用 JavaScript 脚本、图片和 Flash 动画等作为网站导航；注意内部链接的优化，提高网站内部链接的紧密性，提高页面之间关键词的相关度。

(8) 外部链接优化：选取高权重、同行业的网站建立外部链接，慎重交换链接，定期检查外部链接。

(9) 图片优化：通过 img 的 alt 属性设置图片替换文本；通过 URL 链接来提高图片与关键词的相关度；通过文件名、URL 文件夹名称进行图片 SEO；环绕图片的文字对搜索引擎蜘蛛可能造成干扰。

(10) JavaScript 与 Flash 处理：使用 js 文件封装 JavaScript 代码，调换 HTML 代码中 JavaScript 的位置，尽量将 JS 代码放在 HTML 代码之后。尽量避免使用 Flash，制作与 Flash 中内容相同的文字网页。

以上 10 种技术 (其中，每种技术的简要说明计 1 分)；答对 5 种者可