

一种多感官交错式搜索引擎的设计

摘要：多感官交错式搜索引擎（Multi-sensory search engine），是指一种可实现文字、图片、音频、视频、气味等不同类型的信息相互进行搜索的新型搜索引擎。在这样一个信息爆炸的年代，这种搜索引擎可以有效的提高数据的利用效率，帮助人们更准确的获得自己想要的类型的信息，使搜索更加智能化、人性化。本文将简要阐述该搜索引擎的创意来源，基本原理，技术难点，相比于现有搜索引擎所具备的优势以及推广前景。

关键词：多感官 搜索引擎 网状架构 人性化

0.引言

0.1 创意来源

近年来，随着互联网技术在全球范围内的迅速普及，Internet 上的信息量正在以惊人的速度增长。这些信息以文本、图片、视频、音频等形式存在于网络，使得信息多元化，方便了不同人的各种不同需求。

但是，现今搜索引擎多为建立在文本之上的“单向”搜索，可由文本搜索出其他文本或者图像、视频、音频等，而反过来却很难做到。比如某同学听到一首好听的歌却不知道名字，就很难通过现有的搜索引擎获得想要的信息。而各种信息越多，搜索的这种局限性也就越明显。

因此，我设想出了一种“网状架构”的新型搜索引擎，命名为“多感官交错式搜索引擎”，它可以克服现有引擎的“搜索单向性”，使得网络搜索更加多元化、人性化，以适应信息爆炸时代用户对搜索提出的更高的要求，进而开启搜索技术的新纪元。

0.2 现有相关技术介绍及简评

以图搜图：谷歌（Google）、百度（Baidu）等国内外搜索界巨头现在已经可以初步实现“以图搜图”功能，但技术仍不完善，比如同为 A 同学的两张照片，但由于拍摄时的光线、明暗、角度等不同，在搜索时就很难靠其中一张搜索出另一张。

哼唱搜索：即文本搜索音频的逆过程，用户通过麦克风把音频输入（比如哼唱一首歌），搜索得出该音频所对应的文本信息；这项技术目前也并不完善，仅限于音乐领域，并没有扩展到其他领域之中。

现有的新型搜索模式基本上只有以上两种，但由于技术难度较大，这两项新

技术目前仍有较多的缺点，尚未全面推广。因此不难看出，在搜索引擎这个方向上还有很多值得我们去探索的东西，而这一领域的革新也必将带动整个信息技术的发展。

1.核心理念及特点

1.1 传统搜索引擎基本原理简介

传统搜索引擎的基本工作原理通常分为 4 个步骤：

①信息搜集：即通过蜘蛛程序从已知数据库出发，访问网页、抓取文件并跟踪该网页中的链接，从而访问更多的网页，并将相应信息存入数据库；

②建立索引：将蜘蛛抓取的页面文件分解并以巨大表格形式存入数据库的过程，页面文件的相关信息均有记录；

③搜索词处理：即对搜索词的“修饰”过程，去除无用的部分；

④排序：从索引数据库找出所有包含关键词的网页，并按照排名算法得出顺序，之后以一定格式返回用户界面；

从这个过程不难看出，传统的搜索模式基本是完全建立在文本信息之上的，通过文本搜索文本，图片、音频、视频仅是包含该文本网页的附加品（如下图），这样的模式是线性的，具有“单向性”，已经逐渐难以满足用户的需求。

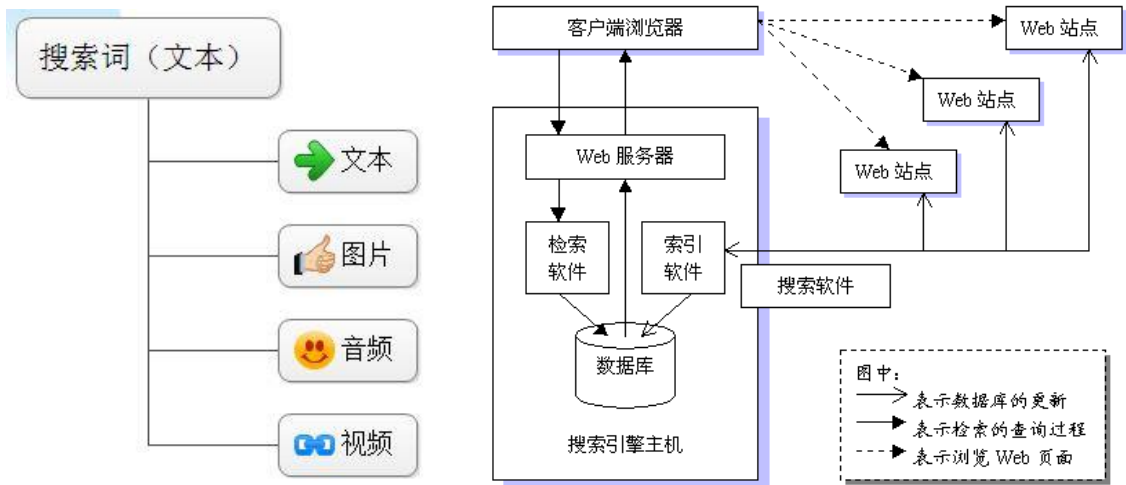


图 1 传统搜索模式图

图 2 传统搜索引擎工作原理图

1.2 多感官交错式的网状搜索新模式

这种新型的搜索引擎设计，主要的创新点在于突破了传统搜索引擎的“一维性”和“单向性”，具有“多核心”、“多方向”、“多感官”的特点，下面对这三个特点逐一进行介绍：

1.2.1 多核心

传统搜索引擎以文本为核心，几乎所以搜索均以文本为出发点，并围绕文本展开；而在这种新式的搜索引擎中，文本、图像、音频、视频均可作为搜索的出发点，从而使引擎本身具备多个核心，搜索可以围绕任何一个核心展开；

1.2.2 多方向

用户可以自主选择搜索的出发点和终点，比如用图片搜索视频、用音频搜索文本等，从而在多个核心之间建立了联系；如果称一对一的搜索确定了一条“搜索线”的话，那么这种多方向的特性便使得“搜索线”之间相互交叉，从而“编织”出一张“搜索网”，极大地提升了搜索的多元性；

1.2.3 多感官

该引擎综合利用了前沿的计算机视觉技术和计算机听觉技术，同时也使用户获得了传统搜索引擎无法提供的全方位的搜索体验；如果未来将“电子鼻”、气敏传感器等计算机嗅觉方面的技术应用于搜索领域，那么用户的体验将会更为丰富，搜索引擎也会更加智能化、人性化。

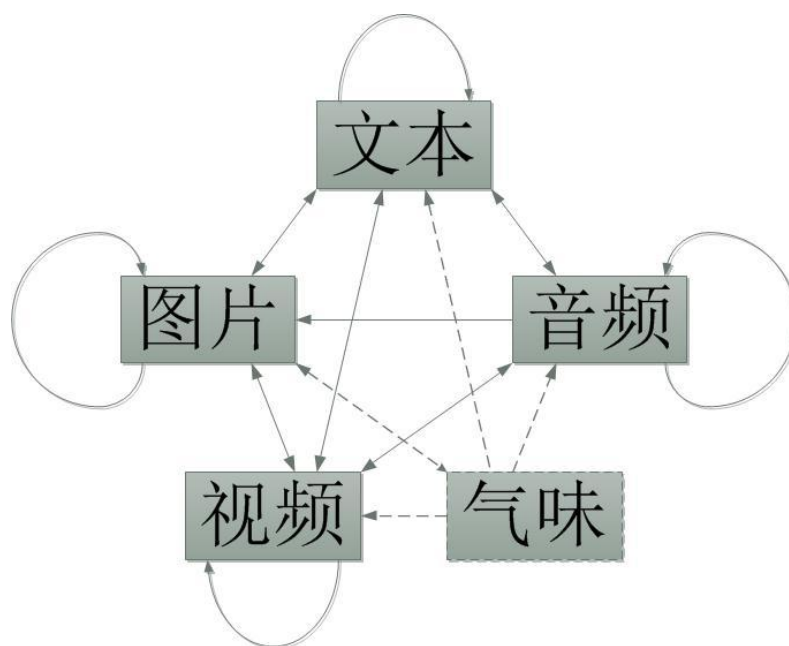


图 3 多感官交错式搜索模式图

2.相关技术分析及设想

由于该搜索引擎运用了许多计算机领域不同分支下的技术，并且部分“搜索线”尚未有相关技术支持，所以下面我将对已有的技术进行简单介绍并尝试提出更优的方法，对尚未成型的技术提出一些合理的设想：

2.1 以文本为核心的搜索

应用传统的搜索方法（参见前文 2.1）基本上可以满足要求，但传统方法的一个弊端就是搜索结果的实时性较差，因此我们可以将搜索与社交网络结合在一起，利用社交网络的实时性来优化搜索得到的信息，从而实现“实时搜索（Real Time Search）”（Facebook 已经开始这方面的尝试）；

2.2 以图片为核心的搜索

现有的“以图搜图”引擎通常是利用“感知哈希算法”（Perceptual hash algorithm）或者更为复杂（识别效果也相应更好）的 pHash 算法和 SIFT 算法，其基本原理就是先将图像灰化，计算灰度平均值后将每个像素的灰度与之比较，

从而得到一个按一定法则生成的字符串，称为图片的“指纹”，通过比较“指纹”判定两张图片的相似度（即计算两字符串的汉明距离），返回给用户相似度较高的图片。

这种方法的最大缺陷就是只能得到构图几乎完全相同的图片，而同一个人不同时候的两张照片就很难实现用一张搜出另一张。对此我们可以利用以下方法将该技术优化：

①得到一张图片的指纹后，在数据库中找到拥有与之相似的指纹的图片（称为目标图片）；

②在目标图片所在的网页的文本信息中进行“二次检索”，找出描述图片内容的文本信息，之后再进行一次文本搜索；

③如果用户的目的是通过图片搜索文本，则返回“二次搜索”后得到的文本信息所在网页的 URL；如果是“以图搜图”，则返回二次搜索后包含图片信息的网页的 URL；

这种改进一方面使得以图片为核心的搜索更为人性化，但另一方面却明显的降低了搜索效率，而且如何在文本中准确筛选出描述图片内容的信息也是一大难题，因此要想实现还有很长一段路要走；

至于通过图片搜索视频，则难度更大；需要事先将网页上的视频信息按帧分解，计算出每一帧的“指纹”并储存在数据库中，如果图片本身就是某视频中的一张截图，则找到相似指纹后返回包含该视频的网址即可；如果不是，则只能利用上面设计的“二次搜索”技术，以文本为中介实现视频的搜索；

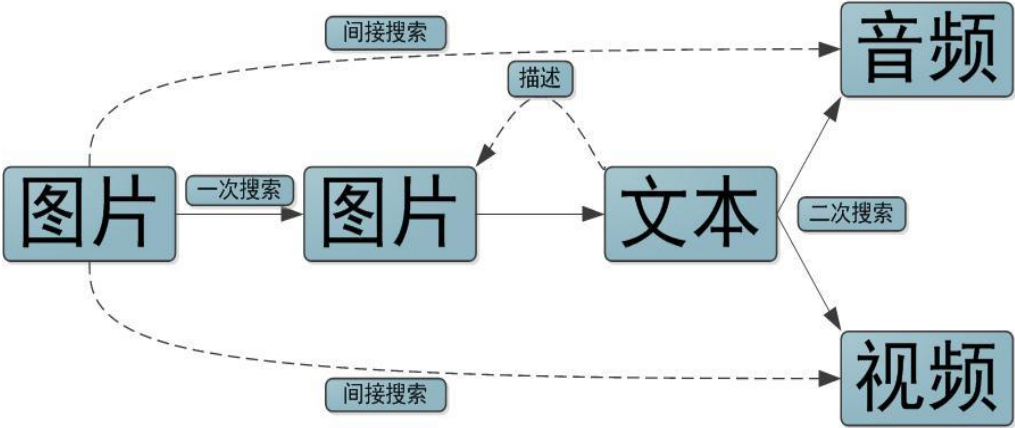


图 4 “二次搜索”模式图

2.3 以音频为核心的搜索

先简单阐述一下现有“哼唱搜索”的基本原理：哼唱搜索通常用一首歌的音高序列作为旋律特征，进而通过中心消波自相关提取法（Center clipping autocorrelation method）获得基频信息，并将之与乐曲数据库中的基频信息相比较，从而获得相应的歌曲信息并返回给用户。至于哼唱时易出现的多音或少音现象，可以通过动态时间规划算法（DTW）解决。

对于歌曲以外的音频信息，我们可以通过“语音搜索”技术实现搜索：这项技术针对用户输入的音频信息，利用一阶隐马尔科夫模型（first-order markov

model) 进行建模，然后以一定法则将其转化为相应字符串，进而找出对应的文本内容，最后进行以文本为核心的搜索。

以上述两项技术为核心，我们可以借助与 3.3.2 中相似的“二次搜索”技术实现“音频搜音频”、“音频搜图片”和“音频搜视频”，从而搭建起以音频为核心的搜索框架；

2.4 以视频为核心的搜索

现有技术尚无法实现以视频为核心的搜索，对此我提出如下设想：

①将用户上传的视频信息按每帧分割（假设共有 n 帧），计算出每帧的“指纹”（假设每帧均被转化成 m 个像素，则每个“指纹”的长度均为 m ），从而得到一个 $n \times m$ 的矩阵（记为 A ）；

②将矩阵的每一行“指纹”分别与数据库中的信息比较，与某一帧相似度高的返回给用户，从而实现“视频搜图片”；

③将矩阵中的“指纹”依次与数据库中信息比较，如果发现连续 x 行相似度满足要求（即 x 行的平均汉明距离小于某个事先确定好的精度 ϵ ，且 $x \geq 100$ ），则可以认为找到了相似的视频（矩阵 B ），返回给用户，从而实现“视频搜视频”；

在上述技术的基础上，利用与 3.3.2 中相似的“二次搜索”技术，可以实现“视频搜文本”和“视频搜音频”，同时“搜图片”和“搜视频”的广度也得到大幅提升，将不只限于相似度高的信息，内容有关联的信息均可被返回给用户；

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{\beta 1} & \cdots & b_{\beta m} \end{pmatrix}$$

图 5 用矩阵表征视频信息

（矩阵 A 与矩阵 B 每行的汉明距离记为 R_i , $i=1, 2, \dots, n$ ；如果存在一个 x ($100 \leq x \leq \min\{n, \beta\}$) 使得 $\frac{R_1 + R_2 + \cdots + R_x}{x} \leq \epsilon$ 成立，则认为 A 与 B 代表的是相似的视频， B 可作为 A 的搜索结果。）

2.5 以气味为核心的搜索

由于关于计算机嗅觉的研究仍在进行中，还很不成熟，所以以气味为核心的搜索基本上仍处在空想阶段，大致思路如下：

通过“电子鼻”、气敏传感器等设备输入气味信息，通过某种方法将气味信

息转化为相应的字符串，与数据库中的字符串比对从而获得与之相关的信息。当然这要求数据库中必须已有气味数据的信息，因此“气味搜索”要想成型，也离不开网络技术的发展。（因为需要将气味信息放在网页上以便搜索引擎建立索引）

3. 应用前景

3.1 Internet 上的应用

应用于 Internet 将会是这一搜索引擎的主要发展方向，这项创意将会成为多元的获取信息的工具，提高网络上各种数据资源的利用效率，改善人机关系，使计算机网络真正成为帮助人们便捷获取信息的工具；

3.2 中小型网络中的应用

直接将这项技术应用于 Internet 无疑难度是很大的，因此待技术成熟后我们可以先将其应用于一些中小型网络，比如企业网络、政府网络等。即使这些网络中的数据量相对较小，但对于管理人员来说不借助搜索引擎进行管理仍比较困难。而将这一技术应用于这些中小型网络中，可以帮助管理人员更加方便准确的管理信息，提高工作效率。

4.结束语

多感官交错式搜索引擎中应用的技术虽然现阶段仍有很多不够成熟，但有朝一日当这些技术都足够成熟时，这项创意无疑将会开启搜索引擎的新纪元，人们将能够更加方便快捷的在信息海洋中获取到想要的信息，同时网络时代人们的生活方式也将得以改善，社会将会朝着更为高科技、现代化的方向发展。

[参考文献]

- [1] 吴亚栋. 语音识别基础[R]. 上海交通大学计算机系, 2007-01.
- [2] 袁津生, 李群, 蔡岳. 搜索引擎原理与实践[M]. 北京: 北京邮电大学出版社, 2008.
- [3] 常丽. 基于内容的图像检索技术[J]. 山西科技, 2008 年, 04 期.
- [4] 施晓华, 黄骥. 信息检索新技术应用[A]. 情报科学. 2005 年, 08 期.
- [5] 高大启, 杨根兴. 嗅觉模拟技术综述[A]. 电子学报. 2001 年, 12A 期.
- [6] 鲁帆, 王民. 一个基于哼唱的音乐检索系统[J]. 中国西部科技. 2011 年, 04 期.