

# 信息检索概述

## 所属学科动向

## 情报学科内涵和外延

情报学 intelligence

信息资源管理学科的内涵

## 情报学的定义

**ASIST (美国科学技术情报学会)**：情报学是一门研究情报的行为和属性，以及处理信息使其易于获得和易于使用的最适宜方法的学科。它关注的是与信息的产生、收集、组织、存储、检索、解释、传播、转换和使用相关的知识体。具有跨学科的性质。

**IIS(英国情报科学工作者协会)**：

- ①情报学核心领域，包括信息的生产、采集、评价、组织、存储、传输、检索和传播的理论与实践；
- ②信息管理，各种组织的信息资源管理；
- ③信息技术，包括可用于情报学与信息管理的各种技术。

## 情报学与信息管理学有一定的交叉

从概念角度分析，涉及到数据（事实）、信息、知识、智能（慧）等。

对应于信息管理学的范畴：数据管理、信息管理（处理）、知识管理、智能（慧）管理等。

## 学科外延

隶属于信息科学群（Information Sciences），情报学是一个子集。

信息科学群包括：情报学、信息管理学、图书馆学、计算机科学、信息学与信息论、运筹学、人工智能、决策科学等。

## 情报学科的特性

情报学具有学科独立性。它与图书馆学、计算机科学等学科是并列关系。

情报学具有跨学科特性。它属于一个交叉学科，与信息学、计算机、传播学、管理科学等学科关联。国内外智库与情报学关系密切（涉及科技、经济、国防等多个领域）。

情报学具有明显的社会和人文维度。它超越了技术层面，可归入社会科学范畴。

## 情报学研究热点

6个方面

- ①情报检索/信息检索
- ②数字图书馆（DL）
- ③竞争情报与知识管理
- ④信息构建（IA）
- ⑤数据库知识发现（KDD）
- ⑥信息资源管理

## 信息检索的概念

### 情报的基本概念

“信息检索”是“情报检索”的发展。

**Information Retrieval (IR)**

有代表性的四种表述：

“情报是判断、意志、决心、行动所需要的能指引方向的知识 and 智慧。”

“情报是在特定时间、特定状态下，对特定的人提供的有用知识。”

“情报是作为存储、传递和转换的对象的知识。”

“情报是被传递的知识或事实，是知识的激活。”

情报的基本属性：知识性、传递性、效用性。

**信息**：泛指一切事物运动的状态和运动的方式。（信息是客观的，是情报的源泉，是情报加工的原料）

**知识**：人类社会实践经验的总结，是正确的结论。（知识也是信息的一部分，是情报的重要来源）

**文献**：通常专指具有历史价值的图书文物资料。（文献、资料是情报的主要来源。另一类型：档案）

**数据**：对客观事物的数学表示。经过处理后能对人的决策行为产生直接影响的数据，也是情报。

### 情报检索的基本原理

情报检索：将情报按一定的方式组织和存储起来，并根据用户的需要找出有关情报的过程。

“存”——对大量信息进行的高度组织化的存储；

“取”——面向随机出现的信息需求而进行的高度选择性的检索，且强调快速方便性。

**“存”和“取”二者的关系： \*\***

一是相互依存。没有存储就无从检索，而没有检索其存储将失去意义。

二是相互制约。从存储的角度看，越简单越好，但如果存储过于简单，必然影响检索的质量和效率。即有效的检索是以增加存储的代价为前提。

### 检索与搜索

检索是数据库时代的概念，人们将数据事先存到数据库里，然后根据需要取出来，就是**检索**。搜索是互联网时代的概念，人们将信息资源放在网上，然后将互联网的信息搜罗起来，建立索引，然后进行查找，就是**搜索**。

**情报检索的本质：情报集合与需求集合的匹配与选择。**

**情报集合**就是有关某一领域的文献或数据的集合体，可以向用户提供所需要的知识或事实，或者获取知识的线索。（含选择、采集、组织、存储）

**需求**指用户的需求，往往以用户提出问题或检索课题的形式体现。众多不同形态的需求汇集，就构成了需求集合。（需求往往带有主观性。满足用户需求是检索的目的）

**匹配与选择**是一种机制，负责把需求集合与情报集合进行相似性比较，然后根据一定的标准选出符合需要的情报。（二者联系的桥梁。包括匹配标准和匹配实施者）

**说明：**情报检索中的匹配与选择，可以是手工或机器完成。

进一步讲，情报检索涉及两个问题：

①**情报（信息）**量庞大，即使一个较小的学科领域，文献量也是数以万计。

要实现有效的匹配与选择，需要原始情报的收集、加工使之有序化和特征化（如分类、标引）；原始情报需要压缩（如编目、做文摘）。——组织与存储

②**对用户需求**，需要进行分析加工。

分析需求的内容，提取出主题概念或其它属性，并利用与情报集合相同的标识系统（检索语言）来表示。分析加工之后的情报需求称为提问（Query）。——用户需求表达

于是，原先的情报集合与需求集合的匹配，就简化为用户提问与有序的、特征化表示的情报集合之间的匹配比较。

## 计算机情报检索的概念

计算机的特点：存储量大，速度快，运算准，可靠性高。

计算机情报检索：在人和计算机的共同作用下完成情报的存取操作，从机器存储的大量情报（信息）中自动筛选出用户所需要的情报。

## 计算机情报检索的内涵剖析

情报检索的本质未变，变的是存储方式和匹配方法。

情报用字符串表示；存储用计算机存储器。

匹配由原人工匹配选择变为机器自动比较和运算。

检索过程就是表示用户需求的字符串（情报需求，提问）与计算机内存储的大量字符串（情报集合，数据库）的比较和运算的过程。

计算机情报检索的实现，使检索服务进入了各个办公室和千家万户，促进了各行各业的信息管理现代化。

## 信息检索的含义

# 广义的信息检索

指信息的存储与检索（2个环节）

存储（Storage）——将大量信息有序化组织并存入数据（信息集合）；

检索（Retrieval）——从数据库中选择、查寻用户所需的信息。

存储是检索的基础，检索是存储的反过程。

## 狭义的信息检索

仅指信息检索环节本身。

对用户而言，狭义的信息检索更重要，因为用户需要的是方便、快捷、高效地获取信息内容，而不必了解信息的搜集、加工和存储。

信息检索的本质：对信息集合与需求集合的匹配与选择。

注意：目前提到的信息检索，一般指计算机化的信息检索。

# 信息检索的研究范围与对象

## 研究范围

信息检索是一门研究信息的处理、存储、查询和获取活动与过程的一般规律的学科。它融合了现代计算机技术、通信技术、存储技术和网络技术，是一个新型交叉学科。

一切与信息存储和检索有关的理论、方法、过程、系统及信息集合，都属信息检索的研究范围。

- 可供存储和检索的信息类型，如文献、数据、事实、知识、声音、图形等；
- 各种检索系统及其处理环节，如信息采集、标引、组织、存储、处理、匹配、输出、传送等；
- 各种与信息检索相关的理论、方法和技术。

## 研究对象（内容）

### 信息检索理论研究

具有理论基础支撑与指导作用。

- （1）检索模型。主要是数学模型；
- （2）标引理论。其中手工标引理论包括分类标引和主题标引，自动标引理论包括统计标引、加权标引、引文标引等；
- （3）信息组织理论。包括信息描述理论、知识组织与知识组织系统理论；
- （4）相关性理论。指有关信息与查询相关性判定的匹配标准理论。

### 信息检索方法研究

检索方法是查找信息时所采用的具体方法，对于检索效果和效率有着重要影响。

主要包括布尔检索法、加权检索法、截词检索法、位置检索法、限制检索法、聚类检索法等。检

索方法研究的目的是改进和完善现有的检索方法，同时研制和开发新的检索方法。**检索策略可归入检索方法研究。**

## 信息检索技术研究

检索技术是实现信息检索有效性的手段和保障。

主要包括文本检索技术、图像检索技术、音频检索技术、视频检索技术以及多媒体检索技术。与网络搜索引擎有关的检索技术包括网络信息自动采集技术、网页链接技术、搜索结果排序技术、元搜索技术、网络挖掘技术等。另外，还有智能检索技术、信息检索可视化技术等。检索技术研究的目的是研究开发新的检索技术。

## 信息检索系统研究

是信息检索活动的物质基础。

主要包括系统的结构（含数据资源）、功能、类型、分析、设计、开发、运行、维护、管理及评价。可分为联机检索系统、光盘检索系统、网络检索系统。

## 信息检索语言研究

检索语言是信息检索不可或缺的工具，是用户与系统交流、互动、沟通的媒介，也影响着检索系统的效率。

检索语言包括受控语言（分类语言、主题语言）检索、非受控语言（关键词、自由词、自然语言）检索以及混合语言检索的研究。此外，还有跨语言检索研究、检索语言一体化研究等。<sup>\*\*</sup>

## 信息检索服务研究

服务用户是检索系统的出发点和归宿，必须研究用户。

主要包括用户及需求类型，用户认知、心理、行为分析，各种服务方式和模式的开发，用户满意度评价等。就服务方式而言，通常包括参考咨询服务、定题检索服务、科技查新服务，及延伸服务（如在检索基础上提供行业报告、战略咨询、竞争情报、资信调查等）。

## 信息检索评价研究

目的是改进和完善系统的性能和功能。

通常包括性能评价、效益评价、评价方法与步骤、指标体系等。

## 其它相关领域研究

数字图书馆等。

## 相关领域

## 计算机科学技术

信息检索始终与计算机技术密切相关。现代信息检索的理论、方法与技术几乎都是借助于计算机技术的成果。

主要领域知识基础包括：程序设计、数据结构、数据库原理、信息系统分析与设计、网络技术等。

## 数学

数学是自然科学的基础，也是社会科学、人文科学研究必不可少的工具。研究检索模型、检索算法、系统评价等，都离不开数学工具。

语言学。由于文本检索是信息检索的主体，检索系统处理的信息大部分都是语言信息，涉及语言信息的输入、形式化表示、编辑、存储、检索、输出及展示等，以及语言的分析。

## 人工智能与认知科学

认知科学是研究人类思维和方法的科学，检索系统涉及检索过程中人、交互、反馈等；人工智能研究中的知识表示、自动推理、机器学习、专家系统、自然语言理解等推动信息检索的发展。