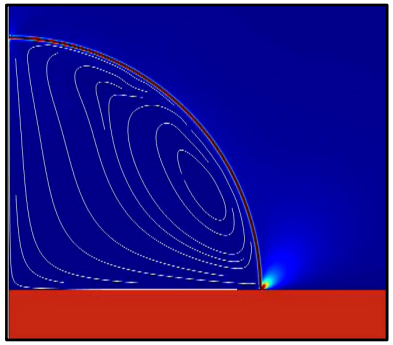
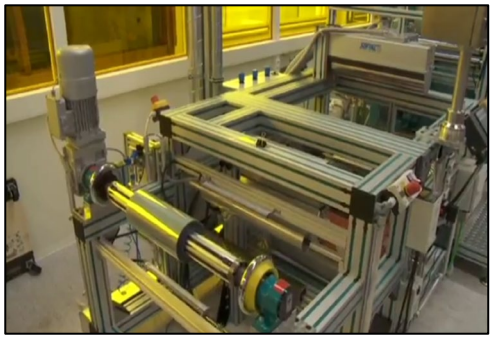
Information Retrieval in Nanoscience and Technology

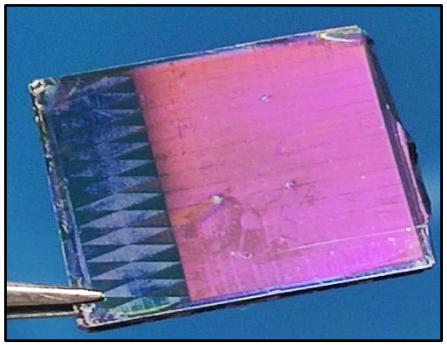
Ass. Prof. Dr. Wei Deng
Institute of Functional Nano & Soft Materials
College of Nano Science & Technology
Soochow University
Email: dengwei@suda.edu.cn

有机印刷电子材料与器件

一、高性能材料的先进印刷制备技术开发





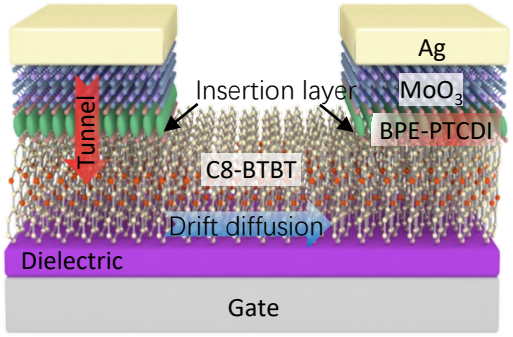


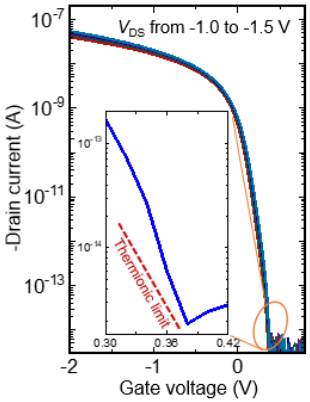
苏大优秀青年学者

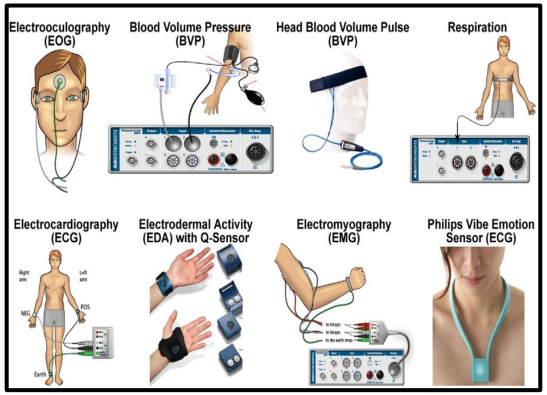
互联网+省赛二等奖

指导教师

二、低功耗、高增益柔性晶体管电路研发







Teaching Assistant Information

FUNSOM

- Graduate student

Name: Yujian Zhang

- E-mail:

851596167@qq.com

- Undergraduate student

Name: Xinwei Zhang

- E-mail:

2628459712@qq.com

- After Class Assignment:

1. Mail: **724119768@qq.com**
2. Assignment naming format: 第*周-本/硕-学号-姓名

Introduction of information retrieval

- Introduction to the Course
- Introduction to Information Retrieval
- Elementary Procedure of Information Retrieval
- Information Retrieval History and Developments

Section 1: Course Introduction

Introduction to the Course

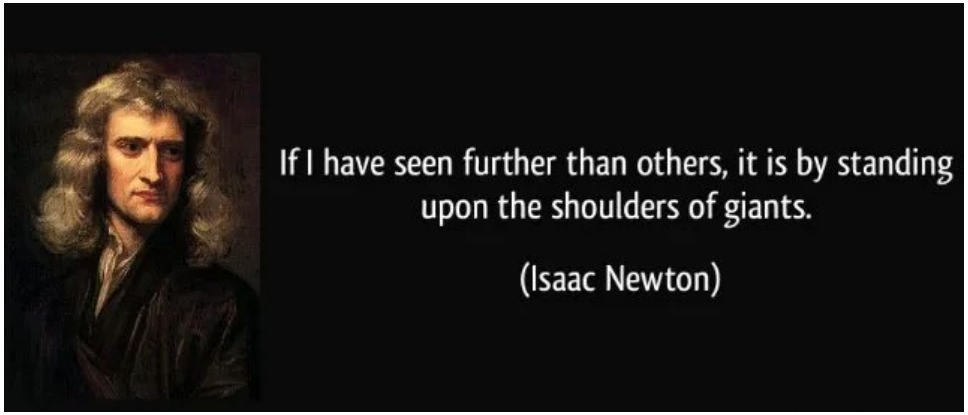
学会

如何从这浩如烟海的信息中

快速且准确地找出所需信息，

主要指科技信息

是我们应具备的一项基本技能。



Introduction to the Course

信息检索顾名思义就是把你想要知道的、了解的信息通过某种途径把它搜索出来。

德国柏林图书馆门前有这样一段话：这里是知识的宝库，你若掌握了它的钥匙，这里的全部知识都是属于你的。这里所说的钥匙即是指信息检索的方法。

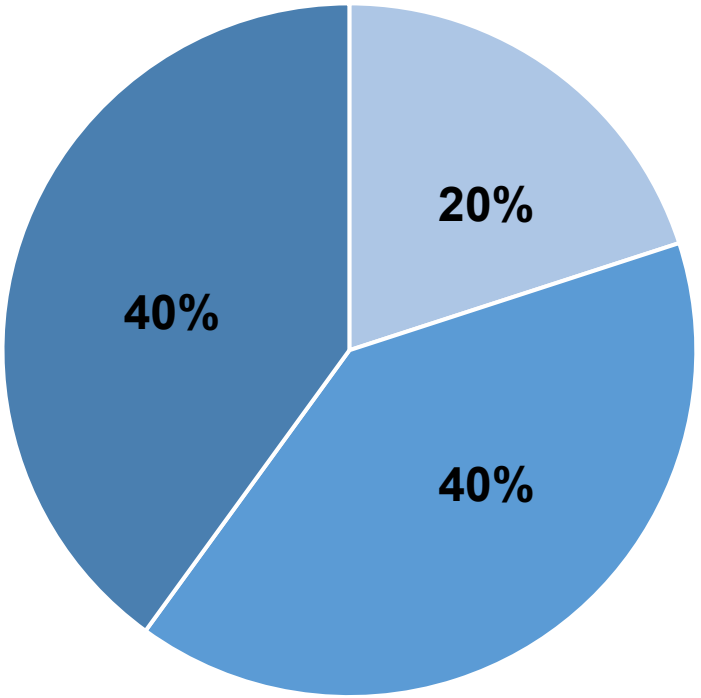
Introduction to the Course

本课程致力于帮助同学们了解世界网络信息资源的类型、特点、分布，认识网络搜索引擎、中外文数据库及其他网络信息资源获取途径，掌握获取、利用、开发信息资源的策略与技巧。

不是一门方法课、技能课，培养学生深入到骨髓的信息检索意识

Grading:

- After Class Assignment: 20%
- Presentation in Class: 40%
- Final Exam: 40%



Course Outline

- **Chapter 1:** Introduction of Modern Information Retrieval
- **Chapter 2:** Fundamental theory of modern literature retrieval
- **Chapter 3:** Technique for literature retrieval
- **Chapter 4:** Literature retrieval strategies and procedures
- **Chapter 5:** Introduction of modern information retrieval platform
- **Chapter 6:** E-journal resources
- **Chapter 7:** Internet resources
- **Chapter 8:** Intellectual Property Right and Patent

Readings

- 时雪峰、龚宏、陈萍秀、刘艳磊，《科技文献信息检索与利用》（第五版），清华大学出版社，2020年
- 马正飞、武文良、王彩凤等，《化学化工信息检索与英语阅读》，化学工业出版社，2015年
- 饶宗政，《现代文献检索与利用》，机械工业出版社，2016年
- Ricardo Baeza-Yates Berthier Ribeiro-Neto, Modern Information Retrieval：The Concepts and Technology behind Search, Addison-Wesley Professional, 2011

Section 2: Information Retrieval Introduction

Introduction Retrieval

What do you think is information retrieval?

Information Retrieval

- Information Retrieval (IR) is finding material of an unstructured nature that satisfies an information need from **within large collections**.

Introduction

- Goal of IR is to retrieve **all** and **only**. The "relevant" documents in a collection for a particular user with a particular need for information.

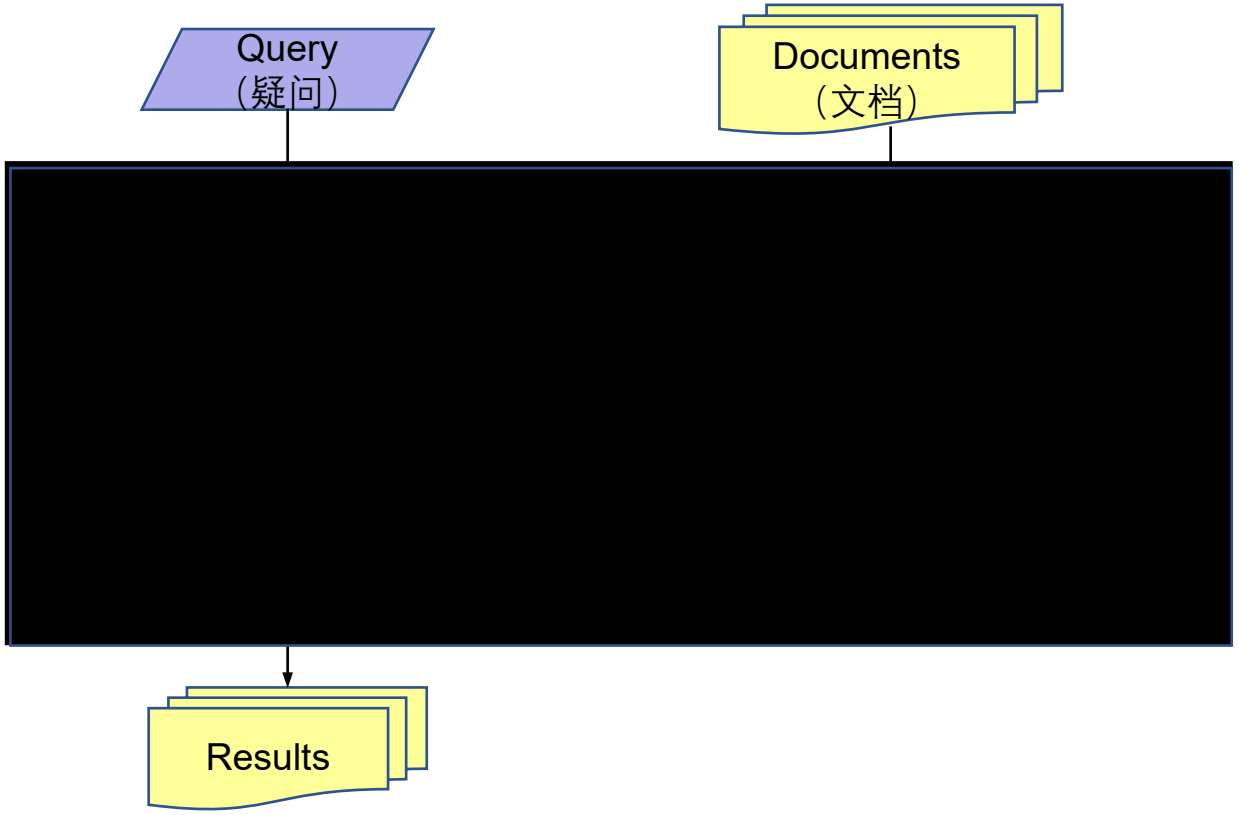
Information Retrieval

- The goal is to search large document collections to retrieve small subsets relevant to the user's information need
- Examples are:
  - Internet search engines (Google, Yahoo! Web search, etc.)
  - Digital library catalogues(MELVYL, GLADYS)

Other IR tasks

- Clustering: Given a set of docs, group them into clusters based on their contents.
- Classification: Given a set of topics, plus a new doc *D*, decide which topic(s) D belongs to (e.g. spam-nospam).
- Information Extraction: Find all snippets dealing with a given topic (e.g. *company merges*)
- Question Answering: deal with a wide range of question types including: list, definition, How, and Why questions.
- Opinion Mining: Analyze/summarize sentiment in a text (e.g. TripAdvisor)
- All the above, applied to images, video, audio

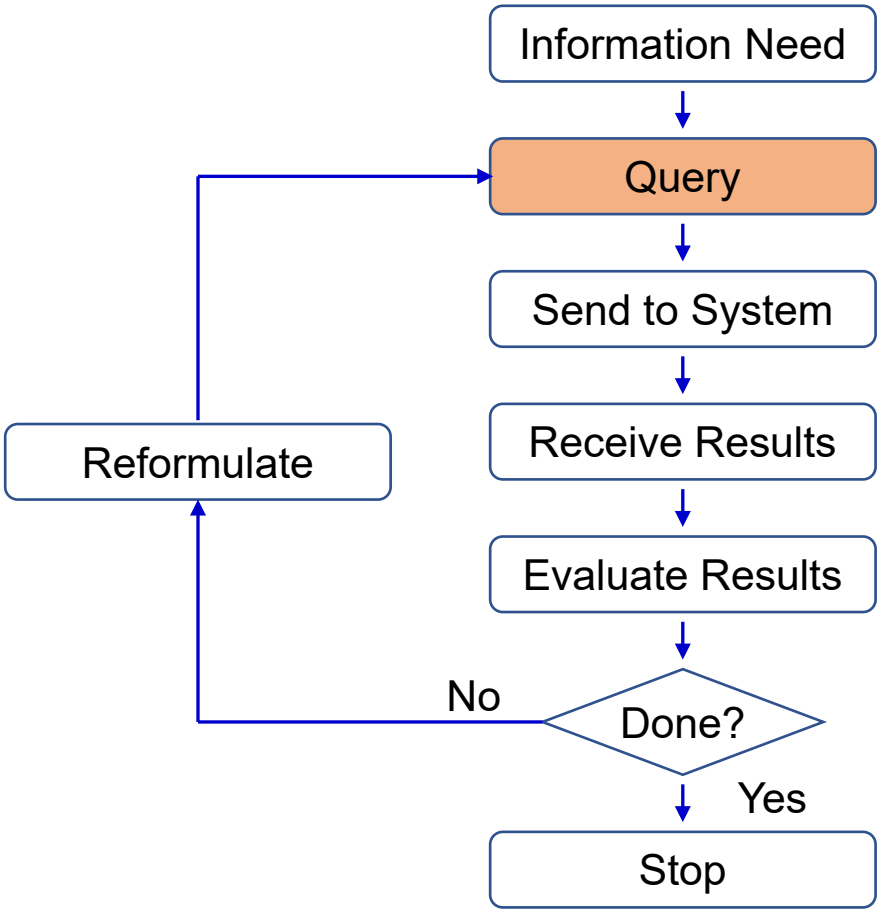
The IR Black Box



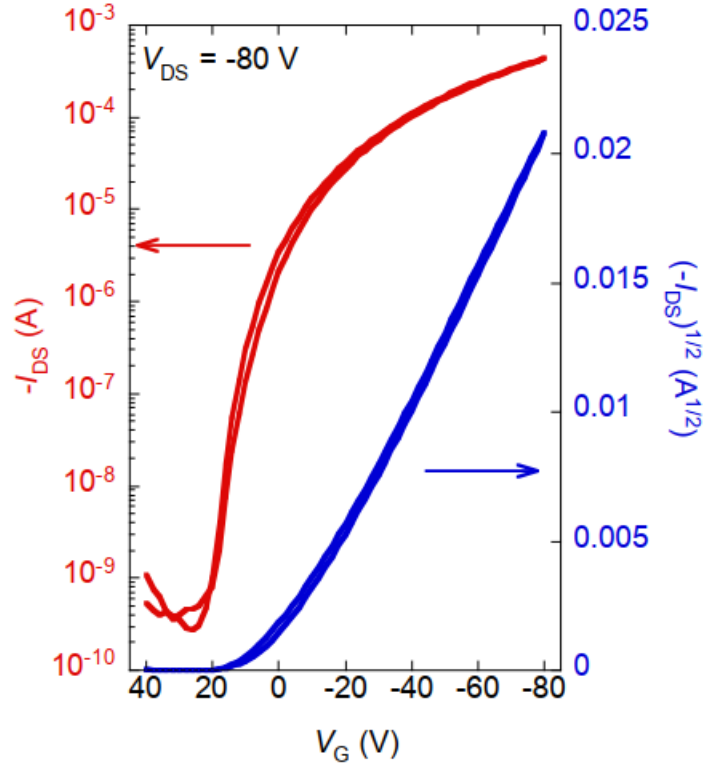
Inside The IR Black Box

Section 3: Elementary Procedure of Information Retrieval

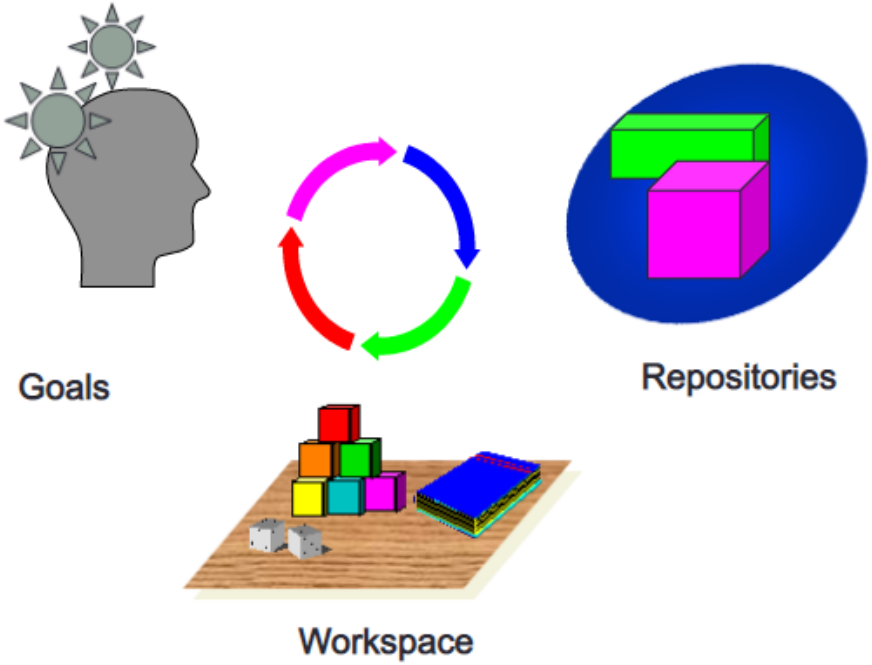
The Standard Retrieval Interaction Model



Example

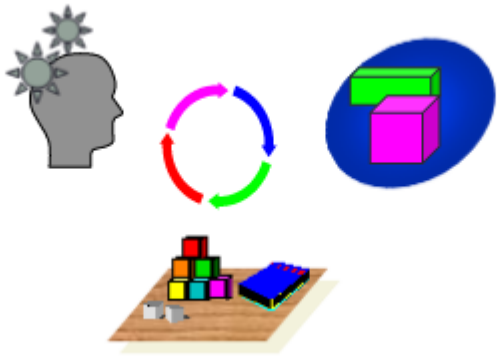


IR is an Iterative Process

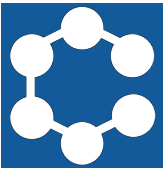


Iteration is the repetition of a process in order to generate an outcome. The sequence will approach some end point or end value. Each repetition of the process is a single iteration, and the outcome of each iteration is then the starting point of the next iteration.

IR is a Dialog



- The exchange doesn't end with first answer.
- Users can recognize elements of a useful answer, even when incomplete.
- Questions and understanding changes as the process continues.



Section 4: Information Retrieval History and Developments

IR History Overview

- Non-Computer IR (mid 1950's)
- Interest in computer-based IR from mid 1950's
- Modern IR-Large-scale evaluations, Web-based search and Search Engines—1990's

信息检索的发展历程

FUNSOM

一、手工检索（1876年）：参考咨询员

二、数字图书馆 / 文档电子化时代（1954年）

美国军队研究机构与高校研究机构的合作，脱机式检索到联机式发展

三、互联网时代（1991年）

四、人工智能时代　内容与行为的精准，弹窗

Computer-Based Systems

- Bagley's 1951 MS thesis from MIT suggested that searching 50 million item records, each containing 30 index terms would take approximately **41,700 hours**