Atitit 人工智能体系树培训列表应用较为广泛的技术.docx

Atitit 人工智能体系培训列表

目录

[1. 1.NLP自然语言处理文本处理 2](#_Toc4361)

[1.1. 语言理解 分词 2](#_Toc3435)

[1.2. 抽取 （压缩文档的读取 格式转换 2](#_Toc25698)

[1.3. 索引处理 摘要提取 2](#_Toc26775)

[1.4. 搜索（按照标题 内容 2](#_Toc14427)

[1.5. 热词检索排行（词云可视化展示 2](#_Toc3289)

[1.6. 专家系统 问答系统 2](#_Toc31021)

[1.7. 智能搜索引擎 数据挖掘和知识发现 2](#_Toc5946)

[1.8. 文本挖掘 2](#_Toc29353)

[1.9. 文本预处理 （网页中存在很多不必要的信息，比如说一些广告，导航栏，html、js代码，注释等等并不需要的信息，可以删除掉 2](#_Toc25475)

[2. 知识图谱 知识处理系统 3](#_Toc23472)

[3. 2.机器视觉 图像处理 3](#_Toc22124)

[3.1. ocr 文字识别 4](#_Toc12836)

[3.2. 条码 二维码识别 4](#_Toc6698)

[3.3. 人脸识别 4](#_Toc5048)

[3.4. 目标识别 验证码识别 4](#_Toc26774)

[3.5. 证件识别 银行卡识别 4](#_Toc25111)

[3.6. 指纹识别 4](#_Toc5942)

[3.7. 图像 视频内容分析 4](#_Toc13603)

[3.8. 图像搜索（相似图片，小图搜大图，人脸搜类似 4](#_Toc31988)

[3.9. 视频编解码五大类 4](#_Toc29797)

[3.10. 其他 图像索引（九宫格模式 4](#_Toc18004)

[4. 3.机器人 在线机器人 自动化 5](#_Toc30960)

[4.1. web自动化 webdriver 5](#_Toc13396)

[4.2. 爬虫（信息采集与信息发布机器人 5](#_Toc25395)

[4.3. gui自动化 5](#_Toc25443)

[5. 生物特征识别(none 5](#_Toc27800)

[6. 人机交互 5](#_Toc25370)

[6.1. 键盘、鼠标、操纵杆 5](#_Toc31458)

[6.2. 打印机、 、显示器、 、音箱等输出设备 5](#_Toc27793)

[7. 4.语言识别 语言tts等 5](#_Toc11169)

[8. Ar vr (none 6](#_Toc13448)

[9. 5.机器学习（决策树 贝叶斯 knn 等 6](#_Toc64)

[10. Other 6](#_Toc32705)

[11. Ref 6](#_Toc22384)

# 1.NLP自然语言处理文本处理

## 语言理解 分词

## 抽取 （压缩文档的读取 格式转换

信息抽取 （压缩文档的读取，zip rar文档

信息抽取格式转换（word，excel，pdf，html文档，eml文档等

## 索引处理 摘要提取

## 搜索（按照标题 内容

## 热词检索排行（词云可视化展示

## 专家系统 问答系统

## [智能搜索引擎](https://baike.baidu.com/item/%E6%99%BA%E8%83%BD%E6%90%9C%E7%B4%A2%E5%BC%95%E6%93%8E" \t "https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD/_blank) [数据挖掘](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98" \t "https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD/_blank)和知识发现

## 文本挖掘

## 文本预处理 （网页中存在很多不必要的信息，比如说一些广告，导航栏，html、js代码，注释等等并不需要的信息，可以删除掉

指剔除噪声文档以改进挖掘精度，或者在文档数量过多时仅选取一部分样本以提高挖掘效率。

例如网页中存在很多不必要的信息，比如说一些广告，导航栏，html、js代码，注释等等并不需要的信息，可以删除掉。如果是需要正文提取，可以利用标签用途、标签密度判定、数据挖掘思想、视觉网页块分析技术等等策略抽取出正文。

# 知识图谱 知识处理系统

知识图谱本质上是结构化的语义知识库，是一种由节点和边组成的图数据结构，以符号形式描述物理世界中的概念及其相互关系，其基本组成单位是“实体—关系—实体”三元组，以及实体及其相关“属性—值”对。不同实体之间通过关系相互联结，构成网状的知识结构。在知识图谱中，每个节点表示现实世界的“实体”，每条边为实体与实体之间的“关系”。通俗地讲，知识图谱就是把所有不同种类的信息连接在一起而得到的一个关系网络，提供了从“关系”的角度去分析问题的能力。

知识图谱可用于反欺诈、不一致性验证、组团欺诈等公共安全保障领域，需要用到异常分析、静态分析、动态分析等数据挖掘方法。特别地，知识图谱在搜索引擎、可视化展示和精准营销方面有很大的优势，已成为业界的热门工具。但是，知识图谱的发展还有很大的挑战，如数据的噪声问题，即数据本身有错误或者数据存在冗余。随着知识图谱应用的不断深入，还有一系列关键技术需要突破

## 微格式

## RDF、RDFS和OWL数据模型

## 语义网络(Semantic Network) 关联数据 链接数据

**语义网络（semantic network）**是一种以网络格式表达人类知识构造的形式。是[人工智能](https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD/9180" \t "https://baike.baidu.com/item/%E8%AF%AD%E4%B9%89%E7%BD%91%E7%BB%9C/_blank)程序运用的表示方式之一。由[奎林](https://baike.baidu.com/item/%E5%A5%8E%E6%9E%97/2555196" \t "https://baike.baidu.com/item/%E8%AF%AD%E4%B9%89%E7%BD%91%E7%BB%9C/_blank)(J. R. Quillian)于1968年提出。开始是作为人类联想记忆的一个明显公理模型提出，随后在[AI](https://baike.baidu.com/item/AI/25417" \t "https://baike.baidu.com/item/%E8%AF%AD%E4%B9%89%E7%BD%91%E7%BB%9C/_blank)中用于[自然语言理解](https://baike.baidu.com/item/%E8%87%AA%E7%84%B6%E8%AF%AD%E8%A8%80%E7%90%86%E8%A7%A3/8465129" \t "https://baike.baidu.com/item/%E8%AF%AD%E4%B9%89%E7%BD%91%E7%BB%9C/_blank)，表示命题信息

**语义网络（semantic network）**是一种用图来表示知识的结构化方式。在一个语义网络中，信息被表达为一组结点，结点通过一组带标记的有向直线彼此相连，用于表示[结点](https://baike.baidu.com/item/%E7%BB%93%E7%82%B9" \t "https://baike.baidu.com/item/%E8%AF%AD%E4%B9%89%E7%BD%91%E7%BB%9C/_blank)间的关系。 [2]

语义网络（semantic network）是一种以网络格式表达人类知识构造的形式。是人工智能程序运用的表示方式之一

一、语义网络

        语义网络是由Quillian于上世纪60年代提出的知识表达模式，其用相互连接的节点和边来表示知识。节点表示对象、概念，边表示节点之间的关系

   但还有个问题，如何区分概念和对象？即定义Class和Object(也称作Instance, Entity)。如果不能区分，会对我们产生什么影响？举个例子，假如我们有两个语义网络A和B。在A中，熊是哺乳动物的一个实例。在B中，熊是哺乳动物的一个子类。前者是is-a关系，后者是subClassOf关系。这种情况常有发生，我们建模的角度不同，那么同一个事物的表示也可能不同。如果我们不能用一种方法来区别两者，不仅会给我们带来理解上的困难，在进行融合的时候也会造成数据冲突。我们不能说A既是B的一个实例，又是B的一个子类。W3C制定的另外两个标准RDFS/OWL解决了这个问题，如下图。

语义网和链接数据是万维网之父Tim Berners Lee分别在1998年和2006提出的。相对于语义网络，语义网和链接数据倾向于描述****万维网中资源、数据之间****的关系。

****关系我们也称为属性（Property）****，根据是实体和实体之间的关系还是实体和数据值之间的关系分为****对象属性****（Object Property）和****数据属性****（Data Property）

# 2.机器视觉 图像处理

计算机视觉是使用计算机模仿人类视觉系统的科学，让计算机拥有类似人类提取、处理、理解和分析图像以及图像序列的能力

根据解决的问题，计算机视觉可分为计算成像学、图像理解、三维视觉、动态视觉和视频编解码五大类

## ocr 文字识别

## 条码 二维码识别

## 人脸识别

## 目标识别 验证码识别

## 证件识别 银行卡识别

## 指纹识别

## 图像 视频内容分析

## 图像搜索（相似图片，小图搜大图，人脸搜类似

## 视频编解码五大类

## 其他 图像索引（九宫格模式

# 3.机器人 在线机器人 自动化

## web自动化 webdriver

## 爬虫（信息采集与信息发布机器人

## gui自动化

# 生物特征识别(none

# 人机交互

人机交互主要研究人和计算机之间的信息交换，主要包括人到计算机和计算机到人的两部分信息交换，是人工智能领域的重要的外围技术。人机交互是与认知心理学、人机工程学、多媒体技术、虚拟现实技术等密切相关的综合学科。传统的人与计算机之间的信息交换主要依靠交互设备进行，主要包括键盘、鼠标、操纵杆、数据服装、眼动跟踪器、位置跟踪器、数据手套、压力笔等输入设备，以及打印机、绘图仪、显示器、头盔式显示器、音箱等输出设备。人机交互技术除了传统的基本交互和图形交互外，还包括语音交互、情感交互、体感交互及脑机交互等技术。

## 键盘、鼠标、操纵杆

## 打印机、 、显示器、 、音箱等输出设备

# 4.语言识别 语言tts等

# Ar vr (none

# 5.机器学习（决策树 贝叶斯 knn 等

# Other

**人工智能技术包含了机器学习、知识图谱、自然语言处理、人机交互、计算机视觉、生物特征识别、AR/VR七个关键技术**

# Ref

人工智能技术包含七个关键技术