

王昊第一次周记

一、 论文选题：

题目：基于深度学习的相似法律案例的判定研究

简介：文本相似度计算方法可分为 4 类：基于字符串的方法、基于语料库的方法、基于知识库的方法和基于神经网络的方法。它是自然语言处理领域中是一项基础性研究工作，伴随着自然语言处理的发展相关方法不断演进。好的相似度计算方法可以大大提高信息检索、查询等系统的效率。

本课题旨在设计一个基于语义分析的文本相似度计算算法，综合考虑词汇在文本中权重、位置等信息，实现准确度更高，适用性更广，具有更大的应用价值的文本相似度计算算法，有效解决“一词多义”与“一义多词”的问题。

要求：

- 1.理解并掌握文本相似度计算相关算法。
- 2.根据现有研究问题，提出有效的基于语义分析的文本相似度计算方法。
- 3.在实际数据集上验证算法有效性。

二、过去一周工作：

过去一周，我主要是学习中外论文搜索和下载，并筛选自己所需要的论文。在一周的工作中，我学会如何在知网和 Google Scholar 中搜索并下载论文，并筛选出了一些自己需要的论文。

三、选择的论文：

- 1、周静, 黄都培, 杨小平, 等. 一种基于特征词的法律文本法条分类算法 [C]. 中国计算机用户协会信息系统分会 2012 年信息交流大会. 2012.
- 2、熊小梅, 刘永浪. 基于 LSA 的二次降维法在中文法律案情文本分类中的

应用[J]. 电子测量技术, 2007, 30(10):111-114.

3、Aletras N, Tsarapatsanis D, Preoțiuc-Pietro D, et al. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective[J]. 2016(10).

4、户保田. 基于深度神经网络的文本表示及其应用[D]. 哈尔滨工业大学, 2016.

5、王超. 面向法律案例的短文本相似度计算研究. 南京师范大学, 2017