**毕 业 设 计（论 文）任 务 书**  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

一、学生姓名：陈功               学号：41924199

二、题目:基于自然语言处理的疾病预测知识库构建

三、题目来源：真实   、    自拟  

四、结业方式：设计   、    论文  

五、主要内容：

中医疾病预测规律的探索，是早期获得病理信息、早期诊断、早期治疗的重要途径，因而具有广泛的社会效益和经济效益，对保护人民健康、推动生产力有极其重要的价值，对中华民族乃全人类的健康事业具有重要意义。

  本课题拟根据中医典籍的语言特征，利用深度学习算法，对中医典籍知识首先基于无监督学习快速获取先验知识、提取特征，减少人工干预和对专业知识的依赖，然后利用深度学习方式进行中医典籍知识的自动获取，最后再实现中医知识的粗图谱构建，以(实体，关系，实体)的三元组形式进行知识表示，并基于图数据库Neo4j实现领域知识图形化展示。

六、主要（技术）要求：

1、以《中医预测疾病100法》等中医典籍为学习对象，利用结巴分词和TF-IDF算法抽取中医典籍中的名词性关键词。利用依存句法分析进行关系抽取。

2、利用BiLSTM-CRF模型进行中医典籍的命名实体识别。

3、利用2Att-BiGRU模型进行中医典籍的实体关系抽取。

4、利用Neo4j图数据库来存储中医理论典籍的知识数据，对实体-关系进行可视化展示，完成知识图谱的构建。

七、日程安排：

第1周—第2周 根据任务书查阅相关文献，完成选题报告，搭建运行环境。

第3周—第4周 搜集相关中医书籍，利用结巴分词和TF-IDF算法抽取中医典籍中的名词性关键词并利用依存句法分析进行关系抽取。

第5周—第7周 使用BiLSTM-CRF等模型实现中医典籍的命名实体识别。

第8周—第11周 使用2Att-BiGRU等模型实现中医典籍的实体关系抽取。

第12周—第14周 使用Neo4j图数据库来存储得到的中医理论典籍的知识数据，对实体-关系进行可视化展示，完成知识图谱的构造。

第15周—第16周 修改论文，准备毕业设计论文答辩。

八、主要参考文献和书目：

[1] 王连心,孟庆刚,王志国,赫炎,徐璞.中药知识库设计浅析[J].世界中医药,2011,6(06):535-537.

[2] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv:1508.01991, 2015.

[3] 张德政, 谢永红, 李曼, 等. 基于本体的中医知识图谱构建[J]. 情报工程, 2017, 3(1): 35-42.

[4] Okanohara D, Miyao Y, Tsuruoka Y, et al. Improving the scalability of semi-markov conditional random fields for named entity recognition[C]//Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics. 2006: 465-472.

[5] 王世昆; 李绍滋; 陈彤生. 基于条件随机场的中医命名实体识别. 厦门大学学报: 自然科学版, 2009, 48.3: 359-364.

[6] Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition[J]. arXiv preprint arXiv:1603.01360, 2016.

[7] STRUBELL, Emma, et al. Fast and accurate entity recognition with iterated dilated convolutions. arXiv preprint arXiv:1702.02098, 2017.

[8] ZHOU, Peng, et al. Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers). 2016. p. 207-212.

[9] LIN, Yankai, et al. Neural relation extraction with selective attention over instances. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016. p. 2124-2133.

指导教师签字： 年 月 日

学 生 签 字： 年 月 日

系（所）负责人章： 年 月 日