**为什么可以采用正则表达式的方式对指南标题等内容进行处理？**

hyf：医学指南是一种体例相对固定的文档。针对每一个病种，指南中列举了相关的症状、病因、相似病种，鉴别要点，在标题命名上规律性很强。所以，一个初步的想法是采用正则表达式来处理。你觉得这样的处理有什么不足之处？可否可以提出其他大的处理方法？

hyf：不同年资的医生，不同场景下，使用指南的方式会不同。请提供一组医生使用指南的场景和需求说明。

hyf：使用医学指南时是否涉及到推理？请提供一组推理的例子。这些例子采用什么样的逻辑语言比较合适？

hyf：电子病历领域在实体关系抽取方面有一些比较深入的探索。你可以下载一些经典论文，尝试采用其中的方法来辨别指南中的实体关系。

hyf：已有的医学本体对于指南的知识抽取有什么帮助？

使用自然语言工具对指南文本进行处理后所得结果中，一些概念之间的关系比较松散。已有的医学知识本体提供了结构化的知识，能够补充这些概念之间存在的潜在关系。

希望你对已有的医学本体资源做个调研和总结，并说明这些本体与基于自然语言处理所得结果如何整合。这些本体包括

* + 谭同学论文中提到的这种英文本体
  + linked data 关联数据网络中涉及的医学本体
  + 其他人论文中提到的疾病本体
  + SNOMED所提供的一些免费资源

hyf：可能涉及的技术、工具与平台：

* 自然语言处理技术
* 本体的相关技术（RDF、OWL、Neo4j等等）
* 逻辑推理技术