各位评委老师好，我叫孙相会，指导教师是张天成老师。我的开题题目是基于机器阅读理解的中文问答系统研究与实现。我的汇报提纲如下，我们先来看一下选题背景与意义。

左图是搜索引擎的例子，搜索引擎的原理是对互联网上数以万计的网页文档中的关键词建立索引库，根据用户输入的问题，基于关键词的检索方式，所有页面中包含问题关键词的网页作为搜索结果检索出来，然后利用排序算法，按照相关度由高到低的顺序返回网页文档。

右图是问答系统的例子，我们说搜索引擎这种基于关键词的检索方式缺乏对用户问题语义的理解，而且返回的是排序的网页文档，仍需要再次筛选答案。而问答系统利用自然语言处理技术，对问题深层次的语义分析，通过检索、推理等方法返回精确的答案，不需要再次筛选。

问答系统大致可以分为三大类：

* 基于知识库的问答系统：通过从知识库中查询与问题相关的实体，推理出实体之间的关系，返回答案。缺点在于需要预先构建好大规模的知识库。
* 基于问答对的问答系统：通过从数据库中检索出与用户问题相似的问题以及对应的答案，然后将答案与用户问题匹配，返回最相似的答案。缺点比较明显：可以回答的问题数量有限，数据库中有多少个问题就只能回答多少个答案，极度依赖于数据集，另一方面返回的答案形式单一，不具有多样性。

我的研究方向是基于机器阅读理解的问答系统，我们来看看什么是机器阅读理解。

机器阅读理解任务是给定问题于问题相关的一篇或者多篇文章，机器阅读这些文章后给出问题的答案。按照答案形式的不同，可以分成抽取式和生成式，生成式是指答案是从文章中归纳总结生成的一段文本，抽取式是指答案来源于文章中的某一段连续的文本。如图所示，目前我的研究计划是做抽取式任务。

基于机器阅读理解的问答系统相比于前两个：

1. 不需要预先构建大规模的知识库
2. 不依赖于数据集，同一篇文档可以回答多个问题

机器阅读理解的应用场景如下。

我们接下来看研究内容与方案：

主要研究内容包括两大部分：

* 设计并实现一个端到端形式的机器阅读理解型问答系统
* 对问答系统中的两个关键部分做深入研究，这两个关键部分是文本匹配模块和阅读理解模块

整个系统的流程是：