Week1

本节课从机器阅读理解介绍、其模型框架和 NIp 数据处理基础三方面阐述，对机器阅读理解进行整体概览。本学习笔记聚焦在机器阅读理解介绍上。

通过阅读从文本中抽取信息并理解意义的过程，称之为阅读理解。人的阅读会自动赋予其本人的思维过程，从而解析文本。然当前的机器学习尚未形成像人一样的自行思考过程，而是聚焦在高层次匹配上。

机器阅读理解（MRC）为在给定文章背景 C(context)和问题集 Q（QA）下，通过学习 C 和 Q 的函数，找出模型函数能正确的预测答案 A(answer)。解释来说：通过 C、Q 交互，从书面文字中提取与构造文章语义的过程。

当前运用的成熟的场景有搜索引擎、机器回答和智能客服上。未来可更多的应用在医疗、法律、金融和教育等行业。

MRC 有以下四大应用：

1）完形填空：在给定文章背景 C 中，移除若干关键词 A 后，通过学习 C-A 函数，能正确填充被移除关键词位置的单词或短语；
2）多项选择：在给定文章背景 C，问题集 Q 和一系列对应问答的候选答案 A（A={A1,A2,…An}中，通过学习 C、Q、A 函数，从一系列候选答案 A 中，找出对应问题的正确答案 Ai;
3）答案抽取：在给定文章背景 C 和问题集 Q 中，其中 C 是由其一系列 token 组成（C={t1,t2,…tn}，通过学习 C 和 Q 函数，提取对应问题集的正确答案 A 在 C 中对应的连续序列，即 A={ti,ti+1,..ti+k}(1<=i<=i+k<=n);
4）自由回答：在给定文章背景 C 和问题集中，学习 C 和 Q 的函数，自由生成正确答案 A，同时 A 可以不来自 C，也可以来自 C。

在答案抽取和自由回答中，本课程主要运用到 SQuAD 和 DuReader 中 Robust 数据集。