# 对于智能BI系统项目的初步理解

**1.关于语义缺失的处理：**

通过与用户交互，继续提问来解决。

**2.智能处理方式，下一个提问与上一个提问相关：**

是否可以让系统具有学习能力，以处理流数据的方式来处理自然语言，为与用户交互的记录建立一个单独的半结构化的数据池（这个我也不太清楚，具体是在coursera的Big Data Modeling and Management System课程上得到的一点启发）。

**3.语义模糊的处理：**

是否可以采用聚类的方法，计算不同单词和短语之间的距离来解决。匹配之后再与用户进行交互，以了解是否定位准确，如果准确再将这项规则添加到数据库或对数据库进行修正。设定一个最大尝试次数，超过之后要求用户重新提供不同的信息。

**4.关于解析词性：**

可以用LTP解决（具体还没有了解），或者在python用jieba解决？

是否必须采用受限汉语？受限程度控制在什么范围？

**5.多词一意的处理：**

同语义模糊的处理。

**6.一词多意的处理：**

是否可以将形容词和名词作为一个整体，而非单独进行解析和聚类？

计算同一个单词的不同含义与语义环境（表征环境的的一系列词）的距离然后取最小？如果差别不大则与用户进行交互？

**7.外部知识库的运用：**

是否可能定期更新一些基本数据，如日期、时间，或者创造在网络或本机上抓取数据的基本功能，在需要的时候调用。

是否可以预设一个常识类的数据池？

**8.设计思路：**

是否可以将常识类和学习到的知识存储到数据池，作为自然语言解析的主体系统，然后设计接口，用于连接专业领域知识的数据库。

主体系统在数据结构上的设计思路需要进一步考虑，coursera的Big Data Modeling and Management System课程可以作为设计的参考。

**以上的理解主要是对自然语言本身的理解，并没有考虑与sql语言的转换问题，而且定义单词与单词之间的距离感觉也很困难。由于阅读的文献太少，可能会有很多不切实际的想法，还望批评指正！**