1. **2011-****Automating String Processing in Spreadsheets Using Input-Output Examples**

最基础的原型。

缺陷：

The algorithm presented in that work assumes strings to be a sequence of characters and can only perform concatenation of input substrings and constant strings to generate the desired output string.[源自2012数字提取的**七、相关工作**]

1. 2011-Learning Semantic String Transformations from Examples 添加了多表搜索的能力
2. **2012-Spreadsheet Data Manipulation Using Examples** 对上面两篇文章的总结与归纳(一定要和上面的结合起来看)。
3. **2012-Synthesizing Number Transformations from Input-Output Examples** 实习生作品，在2011原型的基础上，加入的处理数字的能力(加入了几条小公式)。**重点学习对象**，学习如何小幅改进模型！**又因为他引入了处理数字(大于、小于某个值等)，所以可以用于数据清洗！**
4. **A Colorful Approach to Text Processing by Example.** ：引入了带**结构化本文的处理能力**(比如一本书有署名、作者、出版时间、价格等)。
5. (colorful的进阶,不属于数据提取)**2016 Program Synthesis using Natural Language** ——colorful带来了更复杂的操作，可以通过NLP，输入需求，识别需求，完成输出。
6. 2013-A Machine Learning Framework for Programming by Example 从文本特征中学习得到rules，clues的权重，以加速问题求解。

注：不再使用2011年原型中所用的A\*方法搜索解空间，而是通过学习本文特征，根据概率来决定要使用哪种小程序。(**本文着重处理多行input的问题，对于单行输入，使用原型方法更为有效，也就是说本文没有从本质上提升程序性能。 本文中心在加速，可不看**)

1. 从txt或web提取数据 **[PLDI 2014]: FlashExtract: A Framework for Data Extraction by Example**s ，可用于爬虫等领域，可作为毕设内容。