Swiss Re

General Public Release

On the way towards Intelligent Chatbot

Guan Wang July.14th 2018

Table of Contents / Agenda

- Part I : Chatbot for Vertical Domains
- Part II : NLP Tasks and Algorithms
- Part III: Data, Software Engineering and Project Management

Part I: Chatbot for Vertical Domains

- Are you sure about using chatbot?
- Traditional chatbots
- Natural Language Understanding
- Dialogue Manager
- Natural Language Generation
- End to End Chatbot based on Deep Learning

Are you sure about using chatbot?

Interaction with **Single Target** and **Clear Steps/Logic**: Not Fit

Good for :

- Customer Service in Vertical Domain
- Lots of similar questions and inquiries
- Targets are clear/simi-clear, may need guidance

Domain Expertise (Knowledge Graph) and Deep QA Experience (QA History Data)

Advantage of Chatbot:

- Automatically get User Profile
- Instant Read from Vast Relevant Knowledge Base
- Personalized Answer through Multi-round Dialogue

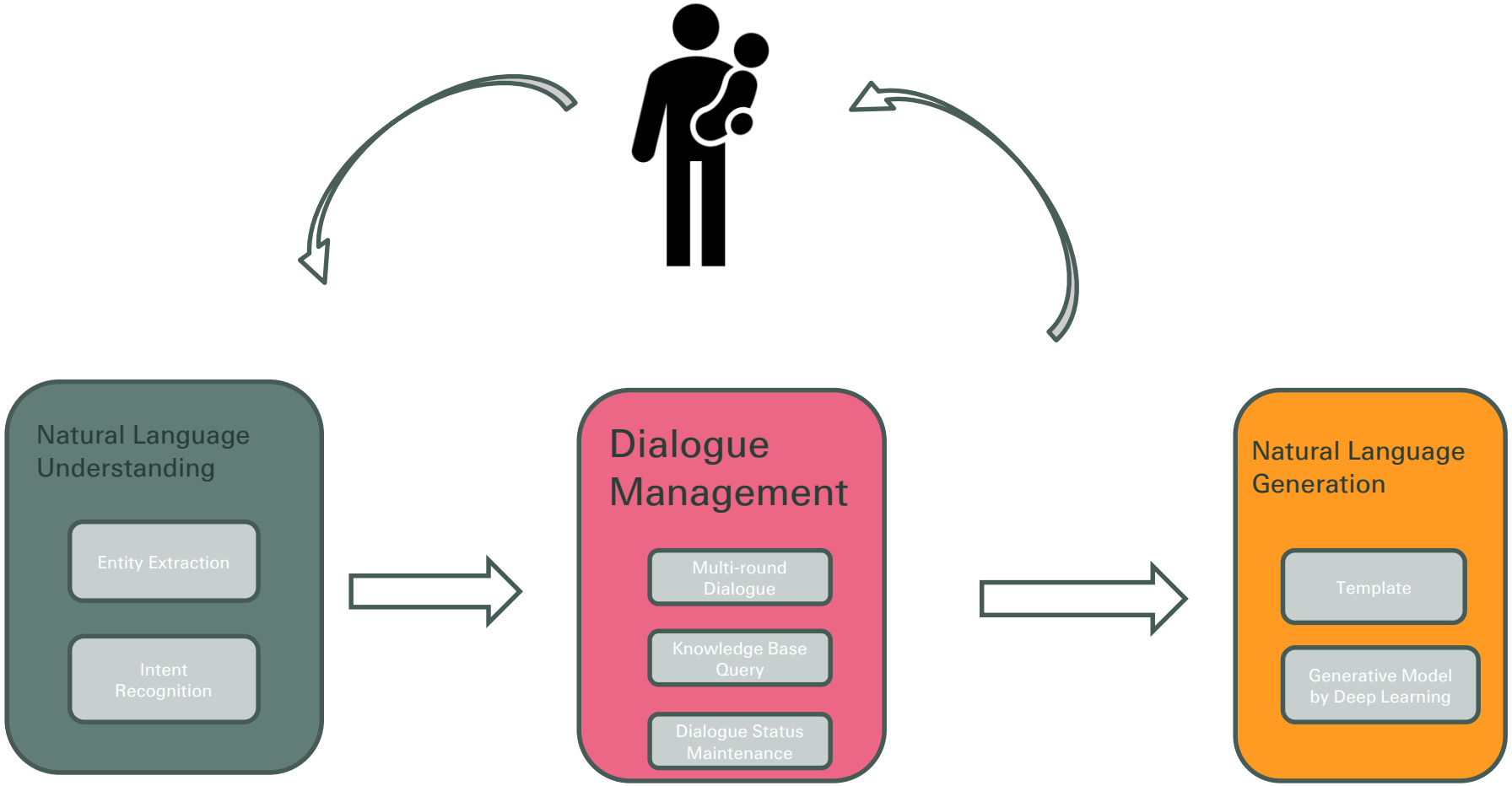
Traditional chatbots

https://github.com/crownpku/aiml_chatbot

- Set of Rules
  - High Accuracy, Low Recall
  - Hundreds of expressions from user for one same intention
  - Difficult to maintain to rule system
- Semantic Similarity
  - Match to Question Database
  - Need large quantity of data
  - Low accuracy

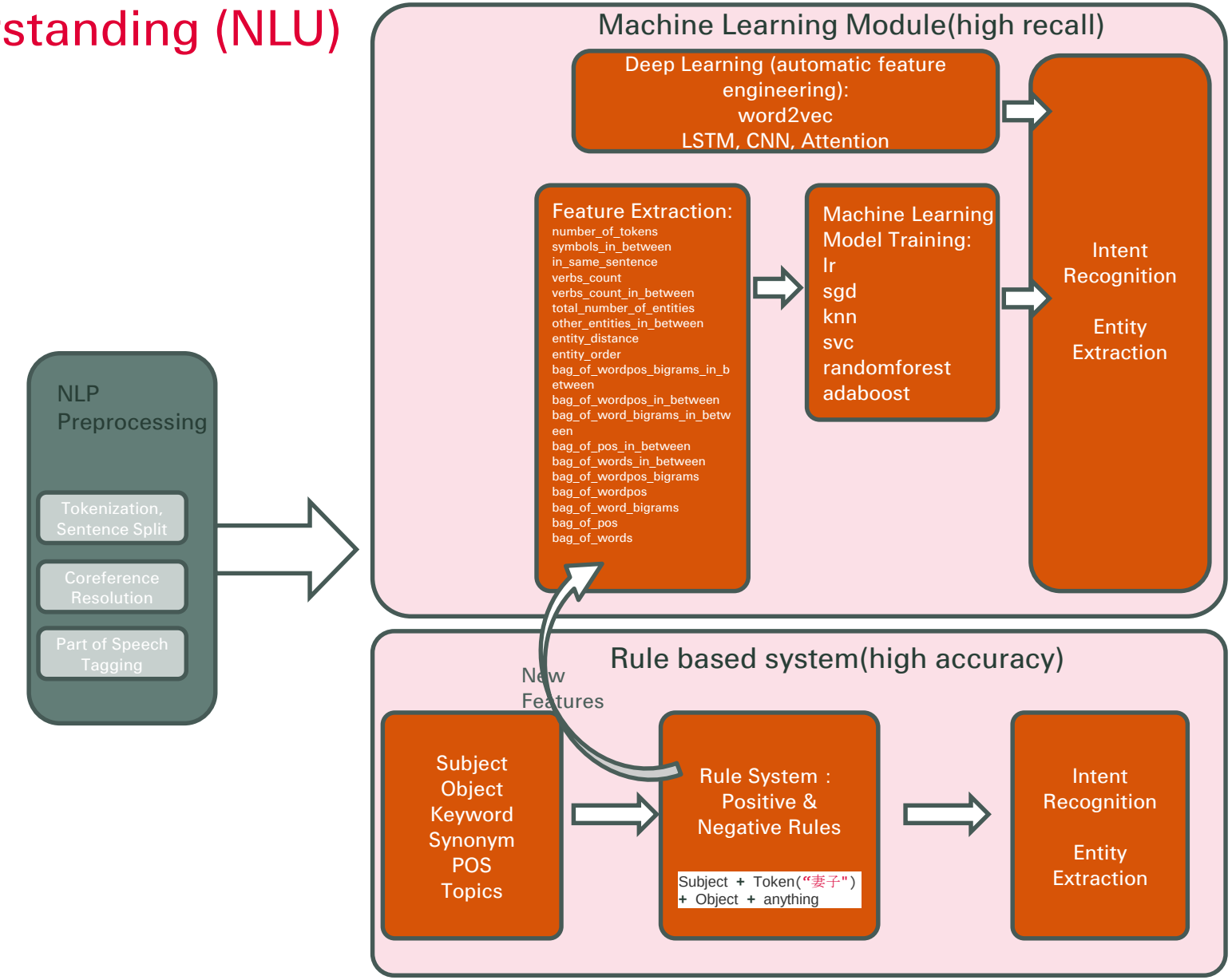

```
<?xml version="1.0" encoding="UTF-8"?>
<aiml version="1.0">

<category>
<pattern>*</pattern>
<that>你现在在什么地方</that>
<template>
<think><set name="where"><formal><star/></formal></set></think>
<random>
  <li><get name="where"/>是个好地方.</li>
  <li>真希望我也在<get name="where"/>, 陪你.</li>
  <li>我刚刚看了下<get name="where"/>的天气哦.</li>
</random>
</template>
</category>

<category>
  <pattern>外面热么</pattern>
  <template>
      你现在在<get name="where"/>,
      <system>python getweather.py realtime <get name="where"/></system>
  </template>
</category>

<category>
<pattern>告诉我 * 天气</pattern>
<template>
<system>python getweather.py realtime <star /></system>
</template>
</category>

<category>
<pattern>* 现在天气</pattern>
<template>
<system>python getweather.py realtime <star /></system>
</template>
</category>

</aiml>
```

Chatbot Architecture

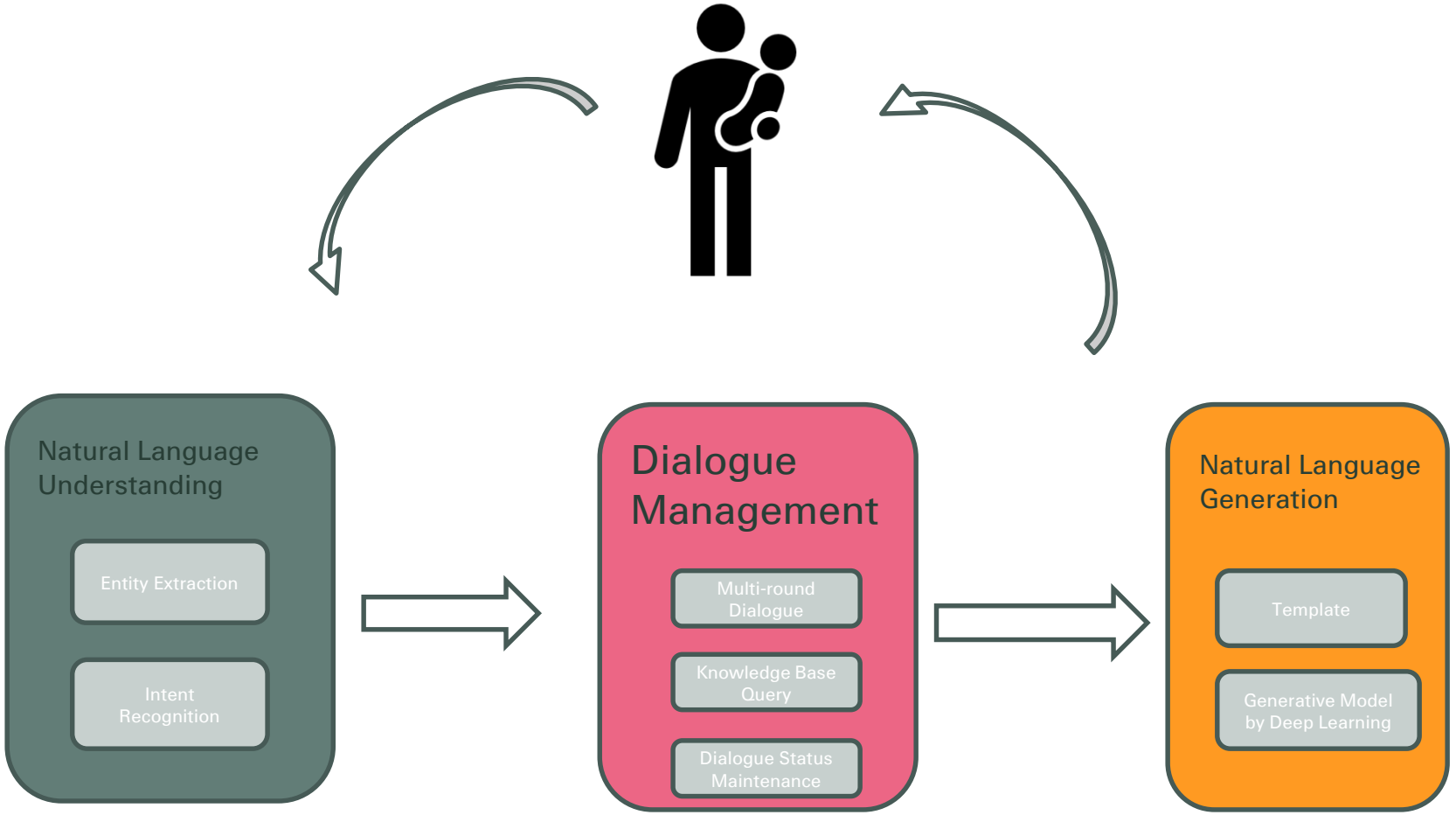


Natural Language Understanding (NLU)


```
{
  "text": "找个吃拉面的店",
  "intent": "restaurant_search",
  "entities": [
    {
      "start": 3,
      "end": 5,
      "value": "拉面",
      "entity": "food"
    }
  ]
},
{
  "text": "这附近哪里有吃麻辣烫的地方",
  "intent": "restaurant_search",
  "entities": [
    {
      "start": 7,
      "end": 10,
      "value": "麻辣烫",
      "entity": "food"
    }
  ]
},
{
  "text": "我胃痛，该吃什么药？",
  "intent": "medical",
  "entities": [
    {
      "start": 1,
      "end": 3,
      "value": "胃痛",
      "entity": "disease"
    }
  ]
}]
```

Dialogue Manager (DM)



User: Kid is sick, what should I do?

NLU Intent Recognition：Sickness
NLU Entity Extraction：Child
DM：No Age, No Symptom, No Gender
NLG

Bot: How old is your child? Boy or girl?
User: 6 months. Boy.

NLU Entity Extraction：0.5 years old; Male
DM：No Symptom
NLG

Bot: What symptom does he have?
User: He is having a fever.

NLU Entity Extraction：fever
DM：Query on Knowledge Base， Define and Finish task
NLG

Bot: Please call Dr. Cai at 13333333333

Dialogue Manager (DM)

- Status Management
  - Current Dialogue Status
  - User History
  - Bot History
  - Possible answers from knowledge base
- Multi-round Dialogue
  - Is user's intent clear?
  - Is entity information enough?
  - Query in a guided conversation
- Execute
  - NLU results into Database Query
    (Knowledge Graph etc.)
  - NLU results into Task Execution
    (Add to shopping cart, Call xx's phone number etc.)
- Implementation
  - Lots of non-standard work regarding to different scenarios
  - Rule based system
  - Supervised Learning as Sequence Labelling problem
  - Deep Reinforement Learning

Natural Language Generation (NLG)

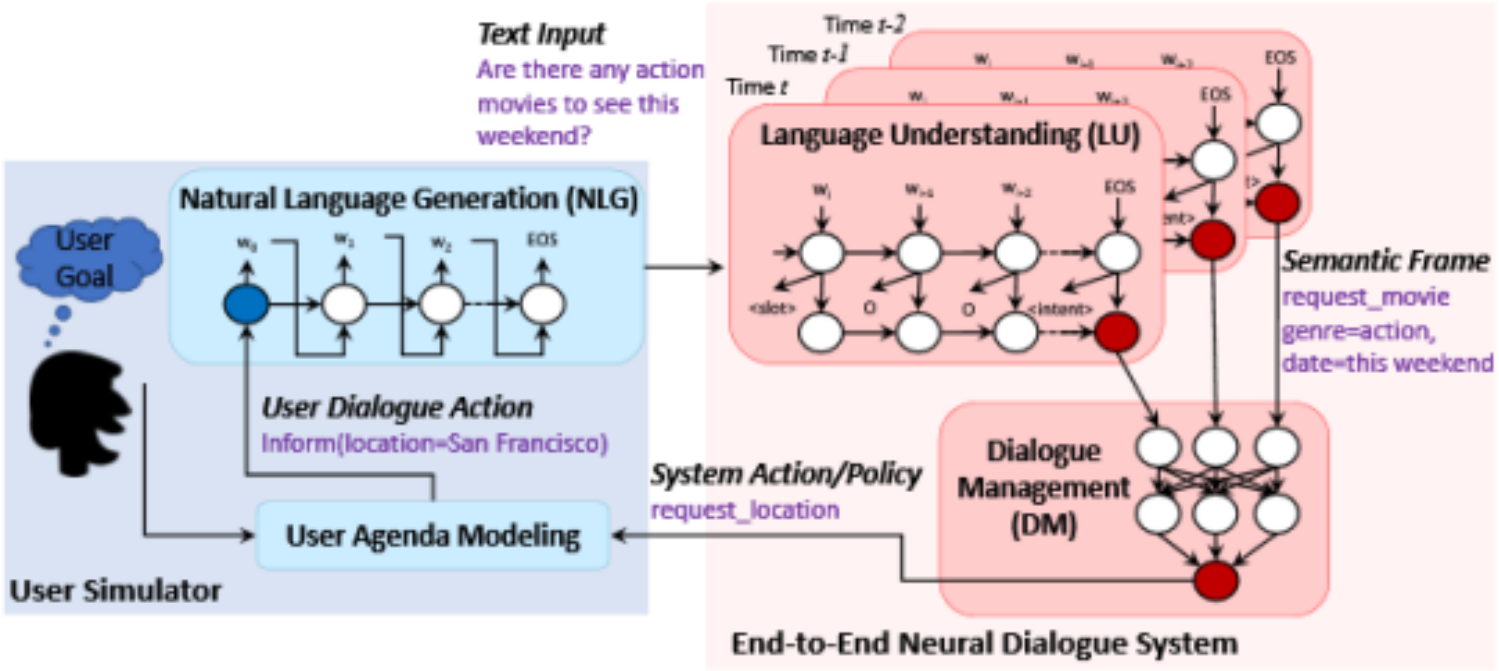
- Template
  - Manually generate template sentences with entity slots
  - Template Filling

- Model
  - Data-driven
  - Learning the templates with entity slots

- Hybrid

End to End Chatbot based on Deep Learning

- NLU+DM+NLG, each module can be DL based



- Memory Network: Embed the whole knowledge base into the model

Part II: NLP Tasks and Algorithms

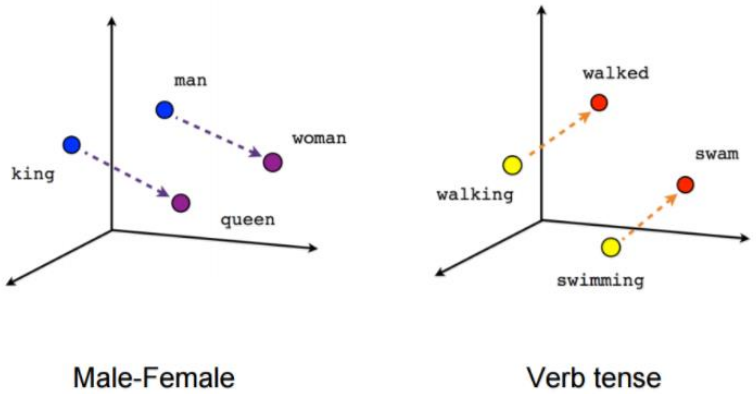
- Overview of NLP Tasks
- Entity Recognition using IDCNN and CRF
- Relation Extraction using BI-GRU and Character Embedding
- Knowledge Graph
- Rasa NLU: Open source Natural Language Understanding
- Somiao-Pinyin: Seq2seq pinyin input method

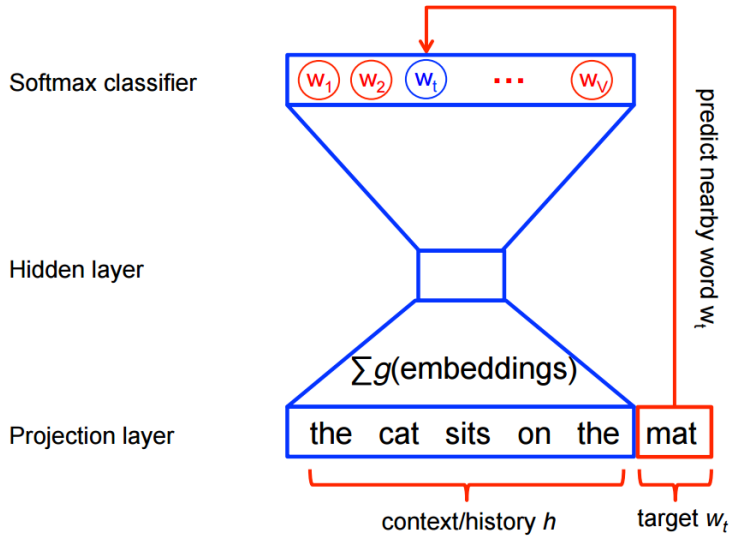
Overview of NLP Tasks

- ❑Word Embedding (text to vector)
- ❑Representation Learning based on Deep Learning
- ❑Large volume of corpus

```
In [14]: for item in y2:
    print item[0], item[1]
   ....:
手枪 0.813866913319
长枪 0.781172335148
猎枪 0.73619812727
散弹枪 0.732991337776
气枪 0.732982099056
土铳 0.713216125965
刀子 0.662224650383
两支 0.655971705914
短枪 0.65476500988
柴刀 0.654682576656
滑膛枪 0.654406666756
霰弹枪 0.648573815823
跳刀 0.635851740837
弹簧刀 0.627648830414
子弹 0.621902227402
枪托 0.621897280216
枪管 0.618635952473
尖刀 0.616757154465
火枪 0.609405577183
一支 0.607889711857
```

```
In [15]: y2 = model.most_similar(u"锤子", topn=20)

In [16]: for item in y2:
    print item[0], item[1]
   ....:
铁锤 0.875185608864
钉锤 0.863791167736
剪刀 0.842311918736
铲子 0.831133902073
斧头 0.829940259457
裁纸刀 0.822092950344
锄头 0.818306088448
镰刀 0.817160069942
小刀 0.813839912415
钳子 0.809854626656
铁棍 0.80567240715
撬棍 0.805454671383
螺丝刀 0.799316167831
尖刀 0.796696186066
铁锹 0.795115530491
钢锯 0.789664387703
```





Overview of NLP Tasks

https://github.com/Kyubyong/nlp_tasks

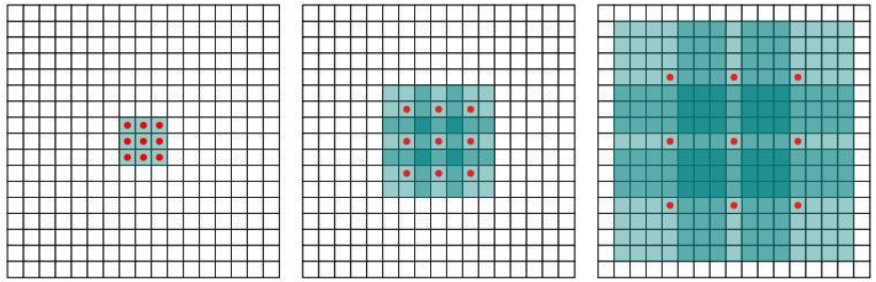
- Category -> Sequence
  - Text Generation, Image Descriptions
- Sequence -> Category
  - Text Classification, Sentiment Analysis
  - Relation extraction
- Sequence -> Sequence (Synchronous)
  - Tokenization, Part of Speech Tagging, Semantic Role Labelling
  - Entity Recognition
- Sequence -> Sequence (Asynchronous)
  - Machine translation, Text summarization
  - Pinyin Input Method

Classification

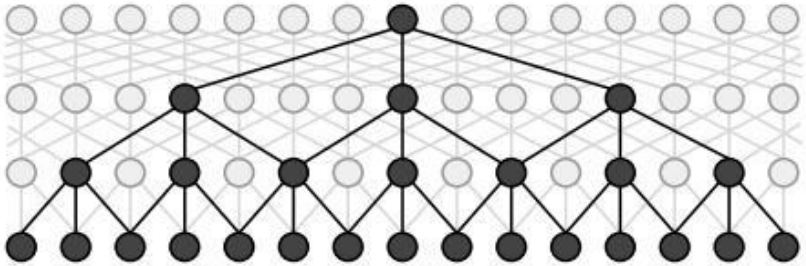
Sequence Labelling

Entity Recognition using IDCNN and CRF

https://github.com/crownpku/Information-Extraction-Chinese/tree/master/NER_IDCNN_CRF



(a) (b) (c)



{'string': '香港的房价已经到达历史巅峰，乌溪沙地铁站上盖由新鸿基地产公司开发的银湖天峰，现在的尺价已经超过一万五千港币。',
'entities': [{'word': '香港', 'end': 2, 'start': 0, 'type': 'LOC'}, {'word': '乌溪沙地铁站', 'end': 20, 'start': 14, 'type': 'LOC'}, {'word': '新鸿基地产公司', 'end': 30, 'start': 23, 'type': 'ORG'}, {'word': '银湖天峰', 'end': 37, 'start': 33, 'type': 'LOC'}]}

{'string': '联想集团的总部位于北京，首席执行官是杨元庆先生',
'entities': [{'end': 4, 'start': 0, 'word': '联想集团', 'type': 'ORG'}, {'end': 11, 'start': 9, 'word': '北京', 'type': 'LOC'}, {'end': 21, 'start': 18, 'word': '杨元庆', 'type': 'PER'}]}

{'string': '在万达集团的老总王健林的著名采访之后，深圳出现了一家公司叫做赚它一个亿网络科技有限公司',
'entities': [{'end': 5, 'start': 1, 'word': '万达集团', 'type': 'ORG'}, {'end': 11, 'start': 8, 'word': '王健林', 'type': 'PER'}, {'end': 21, 'start': 19, 'word': '深圳', 'type': 'LOC'}]}

{'string': '我也不明白为什么有人注册公司名字这么奇葩',
'entities': []}

{'string': '普京和特朗普通了电话，一起表示了对希拉里的鄙视',
'entities': [{'end': 2, 'start': 0, 'word': '普京', 'type': 'PER'}, {'end': 7, 'start': 3, 'word': '特朗普通', 'type': 'PER'}, {'end': 20, 'start': 17, 'word': '希拉里', 'type': 'PER'}]}

{'string': '著名演员刘德华先生，日前在椰椰咕椰岛上拍摄北京遇上西雅图时，从马上摔下受了伤',
'entities': [{'end': 7, 'start': 4, 'word': '刘德华', 'type': 'PER'}, {'end': 18, 'start': 15, 'word': '咕椰岛', 'type': 'LOC'}, {'end': 23, 'start': 21, 'word': '北京', 'type': 'LOC'}, {'end': 28, 'start': 25, 'word': '西雅图', 'type': 'LOC'}]}

{'string': '2015年6月早上发生的那件事，一致停留在李晓华的脑海里',
'entities': [{'end': 24, 'start': 21, 'word': '李晓华', 'type': 'PER'}]}

{'string': '律师解读郭敬明性骚扰事件：若无证据 对李枫不利',
'entities': [{'end': 7, 'start': 4, 'word': '郭敬明', 'type': 'PER'}, {'end': 21, 'start': 19, 'word': '李枫', 'type': 'PER'}]}

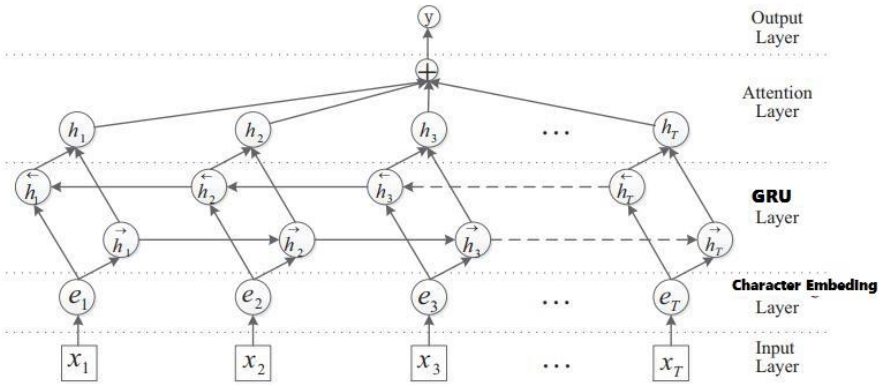
{'string': '南开大学党委书记魏大鹏、校长龚克，中科院院士白以龙、陈和生、陈十一、陈永川、邓小刚、杜江峰、方守贤、葛墨林、贺贤土、洪家兴、江松、李家明、李树深、罗俊、罗民兴、莫毅明、欧阳钟灿、潘建伟、孙昌璞、向涛、谢心澄、邢定钰、杨国桢、张维岩、张伟平、张肇西、赵政国、赵忠贤、周向宇、朱邦芬、邹广田，著名书画家、南开大学终身教授范曾，南开大学副校长严纯华出席。',
'entities': [{'end': 6, 'start': 0, 'word': '南开大学党委', 'type': 'ORG'}, {'end': 11, 'start': 8, 'word': '魏大鹏', 'type': 'PER'}, {'end': 17, 'start': 14, 'word': '龚克，', 'type': 'PER'}, {'end': 20, 'start': 17, 'word': '中科院', 'type': 'ORG'}, {'end': 25, 'start': 22, 'word': '白以龙', 'type': 'PER'}, {'end': 29, 'start': 26, 'word': '陈和生', 'type': 'PER'}, {'end': 33, 'start': 30, 'word': '陈十一', 'type': 'PER'}, {'end': 37, 'start': 34, 'word': '陈永川', 'type': 'PER'}, {'end': 41, 'start': 38, 'word': '邓小刚', 'type': 'PER'}, {'end': 45, 'start': 42, 'word': '杜江峰', 'type': 'PER'}, {'end': 49, 'start': 46, 'word': '方守贤', 'type': 'PER'}, {'end': 53, 'start': 50, 'word': '葛墨林', 'type': 'PER'}, {'end': 57, 'start': 54, 'word': '贺贤土', 'type': 'PER'}, {'end': 61, 'start': 58, 'word': '洪家兴', 'type': 'PER'}, {'end': 64, 'start': 62, 'word': '江松', 'type': 'PER'}, {'end': 68, 'start': 65, 'word': '李家明', 'type': 'PER'}, {'end': 72, 'start': 69, 'word': '李树深', 'type': 'PER'}, {'end': 75, 'start': 73, 'word': '罗俊', 'type': 'PER'}, {'end': 79, 'start': 76, 'word': '罗民兴', 'type': 'PER'}, {'end': 83, 'start': 80, 'word': '莫毅明', 'type': 'PER'}, {'end': 86, 'start': 84, 'word': '欧阳', 'type': 'LOC'}, {'end': 88, 'start': 86, 'word': '钟灿', 'type': 'PER'}, {'end': 92, 'start': 89, 'word': '潘建伟', 'type': 'PER'}, {'end': 96, 'start': 93, 'word': '孙昌璞', 'type': 'PER'}, {'end': 99, 'start': 97, 'word': '向涛', 'type': 'PER'}, {'end': 103, 'start': 100, 'word': '谢心澄', 'type': 'PER'}, {'end': 107, 'start': 104, 'word': '邢定钰', 'type': 'PER'}, {'end': 111, 'start': 108, 'word': '杨国桢', 'type': 'PER'}, {'end': 115, 'start': 112, 'word': '张维岩', 'type': 'PER'}, {'end': 119, 'start': 116, 'word': '张伟平', 'type': 'PER'}, {'end': 123, 'start': 120, 'word': '张肇西', 'type': 'PER'}, {'end': 127, 'start': 124, 'word': '赵政国', 'type': 'PER'}, {'end': 131, 'start': 128, 'word': '赵忠贤', 'type': 'PER'}, {'end': 135, 'start': 132, 'word': '周向宇', 'type': 'PER'}, {'end': 139, 'start': 136, 'word': '朱邦芬', 'type': 'PER'}, {'end': 143, 'start': 140, 'word': '邹广田', 'type': 'PER'}, {'end': 154, 'start': 150, 'word': '南开大学', 'type': 'ORG'}, {'end': 165, 'start': 158, 'word': '范曾,南开大学', 'type': 'ORG'}, {'end': 171, 'start': 168, 'word': '严纯华', 'type': 'PER'}]}

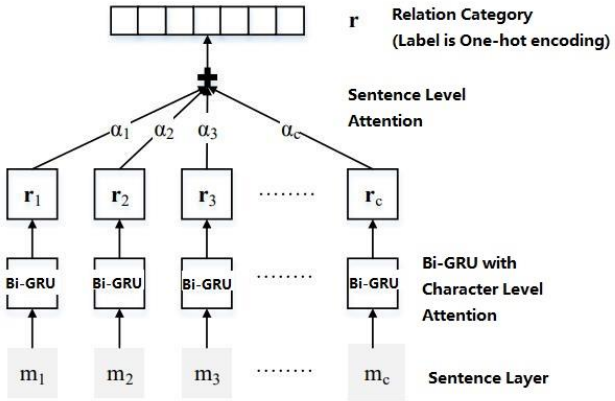
{'string': '陈省身先生的好朋友、原英国皇家学会会长迈克尔•阿蒂亚曾为在爱丁堡广场捐建价值约200万英镑麦克斯韦铜像，花费了很大力气。',
'entities': [{'end': 3, 'start': 0, 'word': '陈省身', 'type': 'PER'}, {'end': 17, 'start': 11, 'word': '英国皇家学会', 'type': 'ORG'}, {'end': 22, 'start': 19, 'word': '迈克尔', 'type': 'PER'}, {'end': 26, 'start': 23, 'word': '阿蒂亚', 'type': 'PER'}, {'end': 34, 'start': 29, 'word': '爱丁堡广场', 'type': 'LOC'}, {'end': 49, 'start': 45, 'word': '麦克斯韦', 'type': 'LOC'}]}

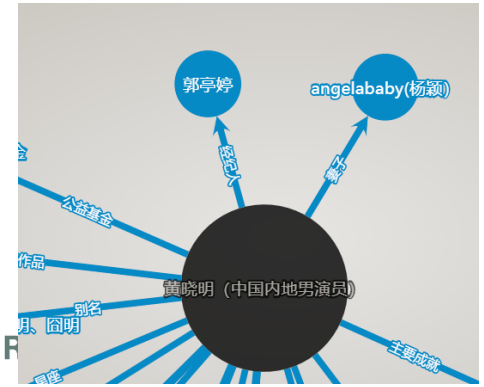
{'string': '当地时间25日(周五)下午2点30分，韩国法院将对三星电子副会长李在镕行贿案作出一审判决。今年49岁、三星集团的实际领导人李在镕，即将迎来他的“命运星期五”。',
'entities': [{'end': 21, 'start': 19, 'word': '韩国', 'type': 'LOC'}, {'end': 29, 'start': 25, 'word': '三星电子', 'type': 'ORG'}, {'end': 35, 'start': 32, 'word': '李在镕', 'type': 'PER'}, {'end': 55, 'start': 51, 'word': '三星集团', 'type': 'ORG'}, {'end': 64, 'start': 61, 'word': '李在镕', 'type': 'PER'}]}

Relation Extraction using BI-GRU and Character Embedding

https://github.com/crownpku/Information-Extraction-Chinese/tree/master/RE_BGRU_2ATT







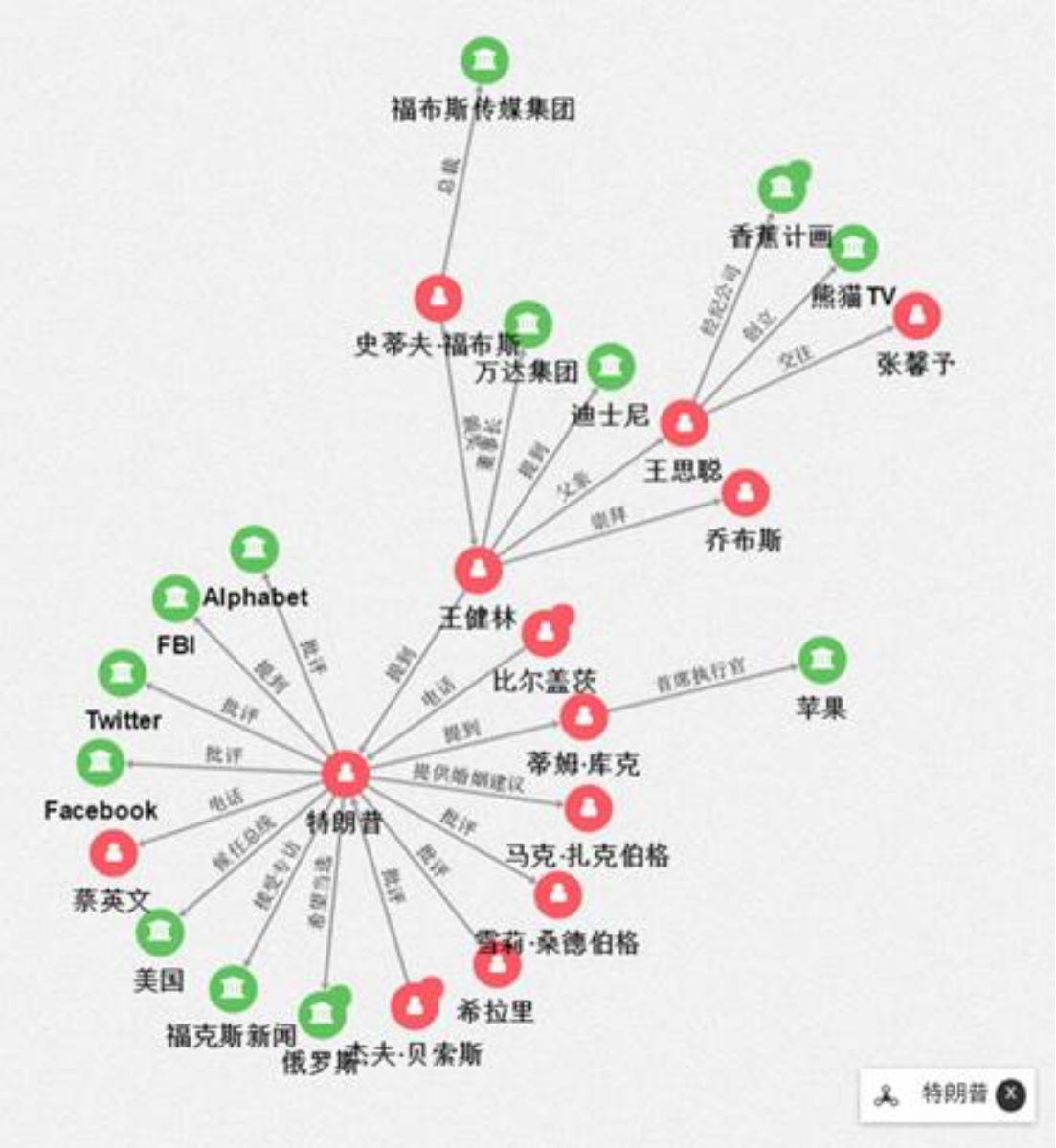

```
实体1： 李晓华
实体2： 王大牛
李晓华和她的丈夫王大牛前日一起去英国旅行了。
关系是：
No.1： 夫妻， Probability is 0.996217
No.2： 父母， Probability is 0.00193673
No.3： 兄弟姐妹， Probability is 0.00128172

实体1： 李晓华
实体2： 王大牛
李晓华和她的高中同学王大牛两个人前日一起去英国旅行。
关系是：
No.1： 好友， Probability is 0.526823
No.2： 兄弟姐妹， Probability is 0.177491
No.3： 夫妻， Probability is 0.132977

实体1： 李晓华
实体2： 王大牛
王大牛命令李晓华在周末前完成这份代码。
关系是：
No.1： 上下级， Probability is 0.965674
No.2： 亲戚， Probability is 0.0185355
No.3： 父母， Probability is 0.00953698

实体1： 李晓华
实体2： 王大牛
王大牛非常疼爱他的孙女李晓华小朋友。
关系是：
No.1： 祖孙， Probability is 0.785542
No.2： 好友， Probability is 0.0829895
No.3： 同门， Probability is 0.0728216

实体1： 李晓华
实体2： 王大牛
谈起曾经一起求学的日子，王大牛非常怀念他的师妹李晓华。
关系是：
No.1： 师生， Probability is 0.735982
No.2： 同门， Probability is 0.159495
No.3： 兄弟姐妹， Probability is 0.0440367

实体1： 李晓华
实体2： 王大牛
王大牛对于他的学生李晓华做出的成果非常骄傲！
关系是：
No.1： 师生， Probability is 0.994964
No.2： 父母， Probability is 0.00460191
No.3： 夫妻， Probability is 0.000108601
```

```
实体1： 李晓华
实体2： 王大牛
王大牛和李晓华是从小一起长大的好哥们
关系是：
No.1： 兄弟姐妹， Probability is 0.852632
No.2： 亲戚， Probability is 0.0477967
No.3： 好友， Probability is 0.0433101

实体1： 李晓华
实体2： 王大牛
王大牛的表舅叫李晓华的二妈为大姐
关系是：
No.1： 亲戚， Probability is 0.766272
No.2： 父母， Probability is 0.162108
No.3： 兄弟姐妹， Probability is 0.0623203

实体1： 李晓华
实体2： 王大牛
这篇论文是王大牛负责编程，李晓华负责写作的。
关系是：
No.1： 合作， Probability is 0.907599
No.2： unknown， Probability is 0.082604
No.3： 上下级， Probability is 0.00730342

实体1： 李晓华
实体2： 王大牛
王大牛和李晓华为谁是论文的第一作者争得头破血流。
关系是：
No.1： 合作， Probability is 0.819008
No.2： 上下级， Probability is 0.116768
No.3： 师生， Probability is 0.0448312
```

Knowledge Graph

- **Construction of Knowledge Graph**
  - **Entity Recognition -> Nodes**
  - **Relation Extraction -> Links**
  - **Articles -> Graphs**

- **Applications based on Knowledge Graph**
  - **Visualization/Exploration**
  - **Graph Based Algorithms**
  - **Graph Database (Relational and NoSQL)**



Rasa NLU: Open source Natural Language Understanding

https://github.com/crownpku/Rasa_NLU_Chi

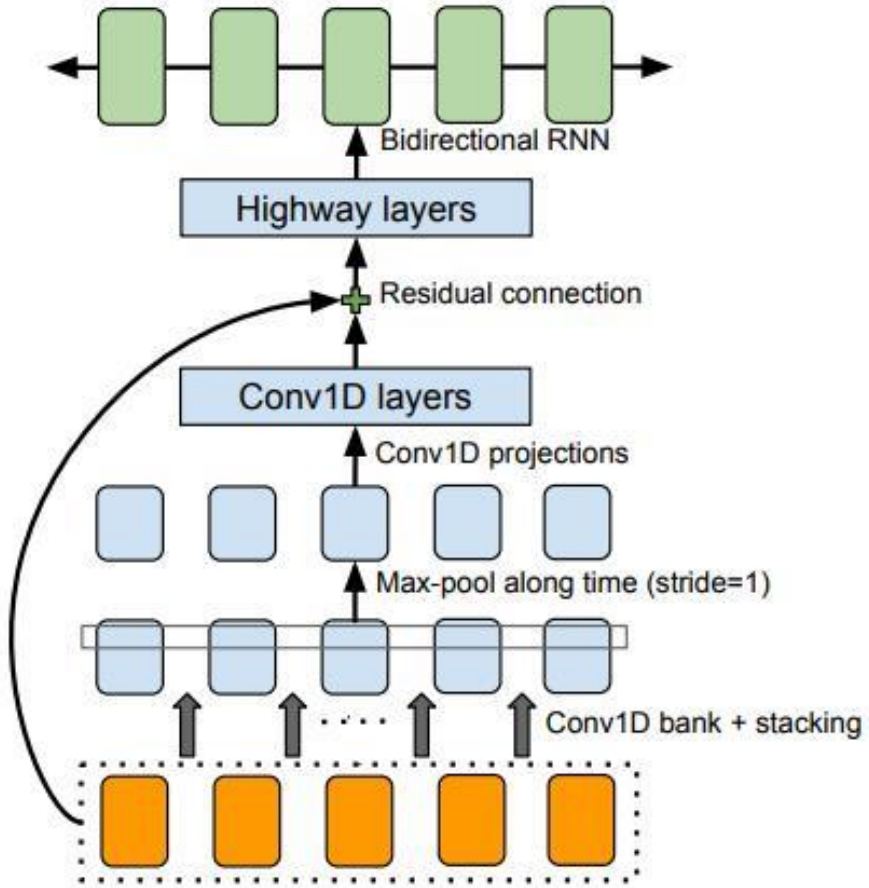

```
{
  "text": "找个吃拉面的店",
  "intent": "restaurant_search",
  "entities": [
    {
      "start": 3,
      "end": 5,
      "value": "拉面",
      "entity": "food"
    }
  ]
},
{
  "text": "这附近哪里有吃麻辣烫的地方",
  "intent": "restaurant_search",
  "entities": [
    {
      "start": 7,
      "end": 10,
      "value": "麻辣烫",
      "entity": "food"
    }
  ]
},
{
  "text": "附近有什么好吃的地方吗",
  "intent": "restaurant_search",
  "entities": []
},
{
  "text": "肚子饿了，推荐一家吃放的地儿呗",
  "intent": "restaurant_search",
  "entities": []
}
```


NLU:
Entity Recognition
Intent Recognition

Online service with Restful API:
API.ai (Google)
LUIS.ai (Microsoft)
WIT.ai (Facebook)
KITT.ai (Baidu)

Somiao-Pinyin: Seq2seq pinyin input method

https://github.com/crownpku/Somiao-Pinyin




```
请输入测试拼音：nihao
你好

请输入测试拼音：chenggongle
成功了

请输入测试拼音：wolegequ
我了个曲

请输入测试拼音：taibangla
太棒啦

请输入测试拼音：dacolehuizenmeyang
打破了会怎么样

请输入测试拼音：pujinghehujintaotongdianhua
普京和胡锦涛通电话

请输入测试拼音：xiangbuqilaishinianqianfashengleshenme
想不起来十年前发生了什么

请输入测试拼音：meiguohongzhawomenzainansilafudedashiguan
美国轰炸我们在南斯拉夫的大事馆

请输入测试拼音：liudehuanageshihouhaonianqing
刘德华那个时候好年轻

请输入测试拼音：shishihouxunlianyixiabilibilideyuliaole
是时候训练一下比例比例的预料了
```

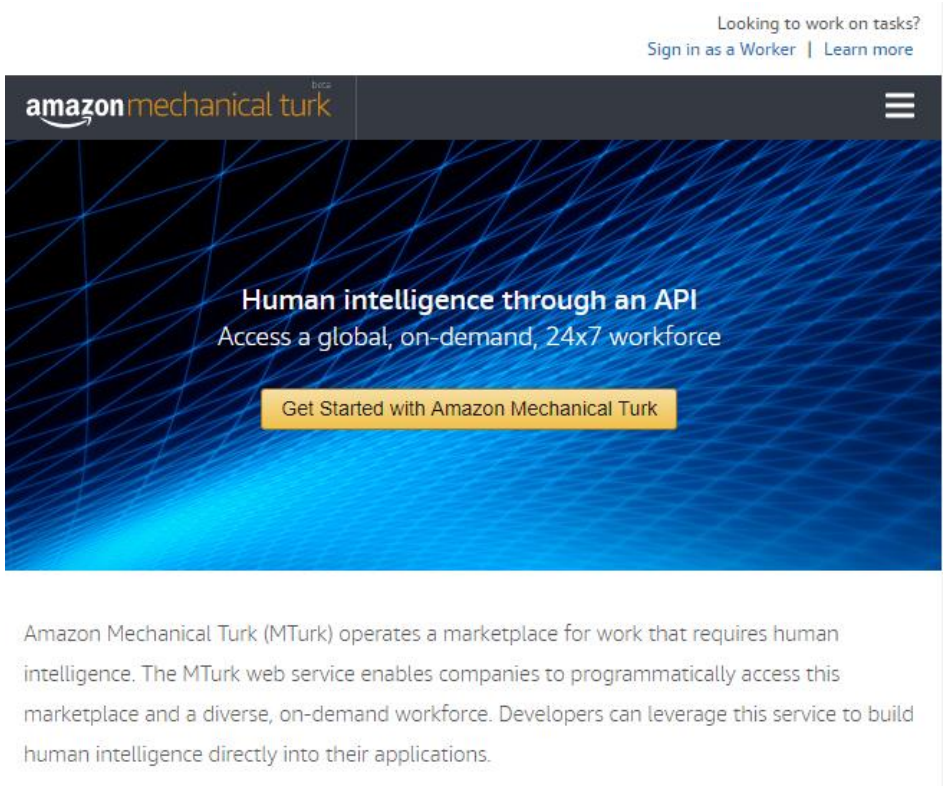
Part III: Data, Software Engineering and Project Management

- Data matters
- Software Engineering matters
- Project Management matters

Data Matters

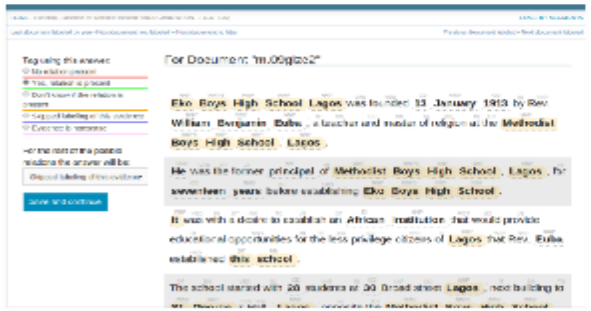
https://github.com/crownpku/Small-Chinese-Corpus
https://github.com/crownpku/Awesome-Chinese-NLP

- Most ML Applications are **Supervised** Learning
- If lucky: automatically generated
  - Click Google ad on a webpage
  - Add stuff in shopping cart on Amazon
  - Listen to music on Spotify
- Not so lucky, but there are smart stricks
  - Crawl from Internet
  - Distant Supervision
- Not lucky, no tricks (usually)
  - Label by hands

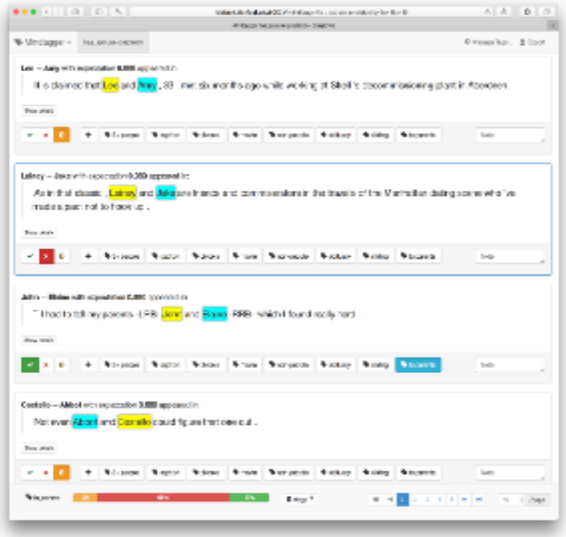


Data Matters: Annotators

IEPY



DeepDive (Mindtagger)



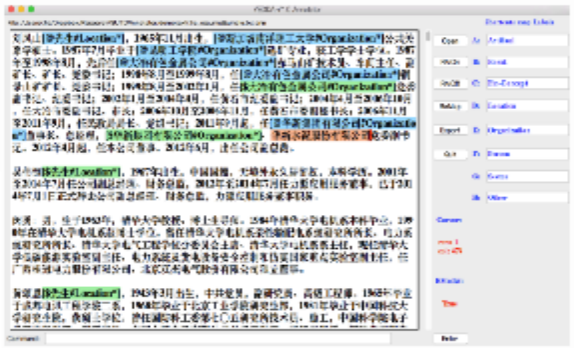
Snorkel

Figure 3: Labeling functions which express pattern-matching, distant supervision, and weak classifier heuristics, respectively, in Snorkel's Jupyter notebook interface.

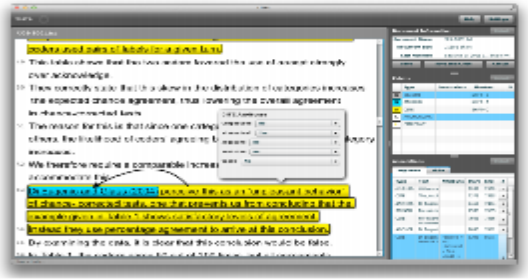
Figure 4: The Viewer utility in Snorkel, showing candidate company-employee relation mentions, comprised of candidate person and company mention pairs.

BRAT

SUTDAnnotator



Slate



Data Matters: Intelligent Labelling with Active Learning

1. User make labels
2. Backend active learning algorithm will consist of "Online" part and "Offline" part.
    - "Online" part will do the online learning and update online model in real time, using fast traditional algorithms like Linear Classifiers or SVM
    - When label data accumulated to a certain amount, "Offline" part will update the offline model, using probably highly accurate deep learning models.
3. After model is updated, predict as much as possible within reasonable time, rank the confidence, and choose the lowest certain number of samples as datasets waiting to be labeled. Repeat step 1.

Software Engineering Matters

https://github.com/crownpku/Chinese-Annotator

Algo Factory

预处理算法:

- 分句、分词
- 句法分析
- 词性标注
- 词库
- 规则库

Online算法:
增量数据
Training/Inference on-the-fly
置信度计算排序
Steam生成代表主数据

Offline算法:
全量数据
准确度优先
置信度计算排序
批量生成待标注数据

协作算法

Database

数据索引模块

同步

数据存储模块:

原始文本存储
Golden标注结果
模型预测结果
批量待标注数据

Task Center

Command line Interface

任务调度 实例管理 逻辑操作
Config读写 数据库读写 模型读写
结果导入导出

Restful API

User Instance

实例config:

Database name
Task Type (NER/RE/Classification)
Label Unit
(分句、分段、分文章、或特定分隔符)
etc...

模型存储:

Online模型参数
Offline模型参数
规则库

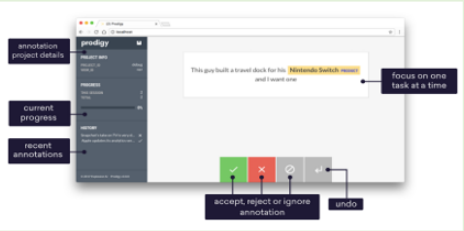
Offline模型config:
Word embedding
LSTM/CNN

Online模型config:

特征提取
number_of_tokens
symbols_in_between
in_same_sentence
verbs_count
verbs_count_in_between
total_number_of_entities
other_entities_in_between
entity_distance
entity_order
bag_of_wordpos_bigrams_in_between
bag_of_wordpos_in_between
bag_of_word_bigrams_in_between
bag_of_pos_in_between
bag_of_words_in_between
bag_of_wordpos_bigrams
bag_of_wordpos
bag_of_word_bigrams
bag_of_pos
bag_of_words

分类模型
lr
sgd
knn
svc
randomforest
adaboost

Web UI



增量标注方式:

命名实体识别-高亮词语选择实体分类
关系提取-高亮两个实体选择关系分类
文本分类-选择文本分类

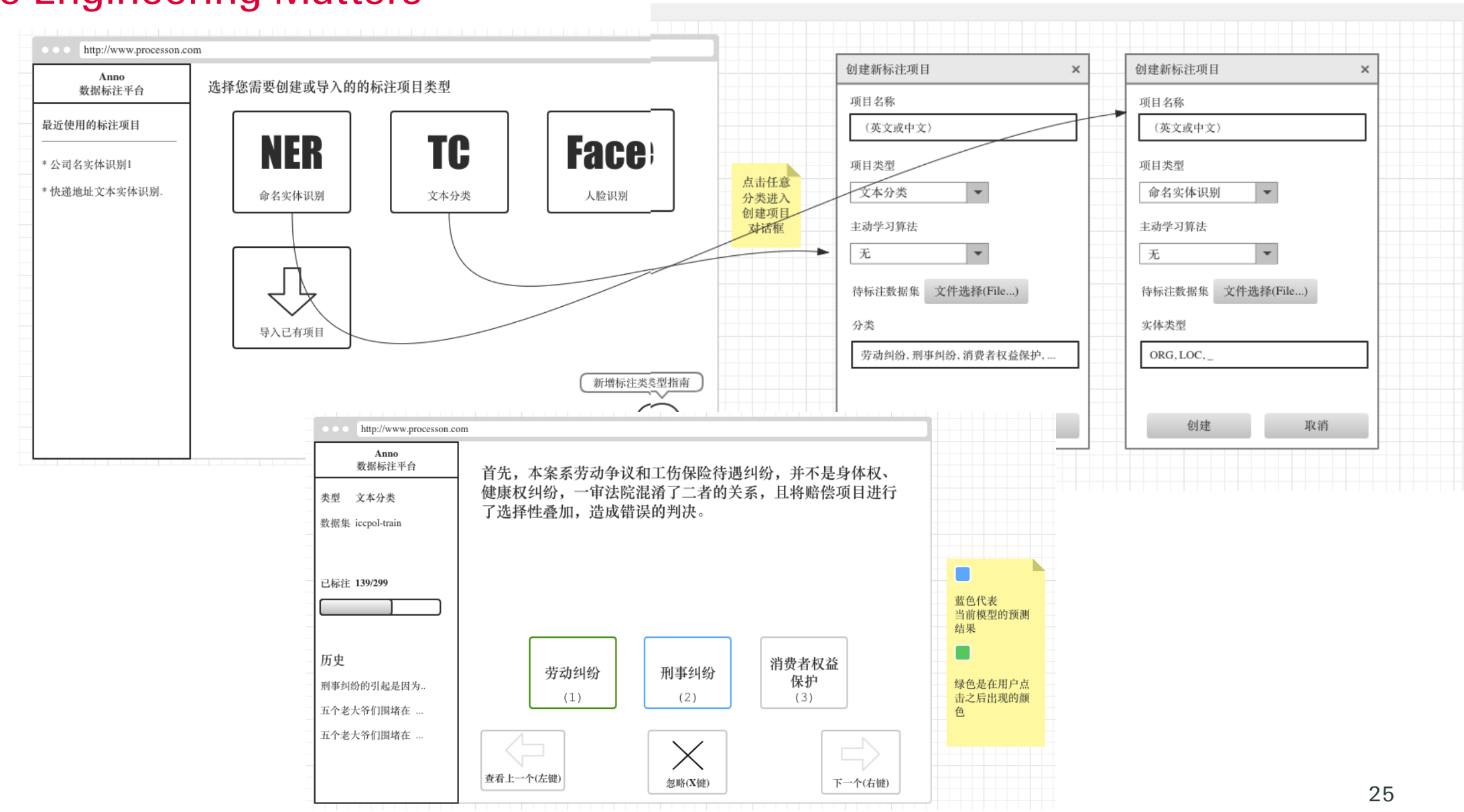
判断标注方式:

无论何种任务，给出模型推断结果
用户点选
正确、错误、忽略、回撤

其它交互:

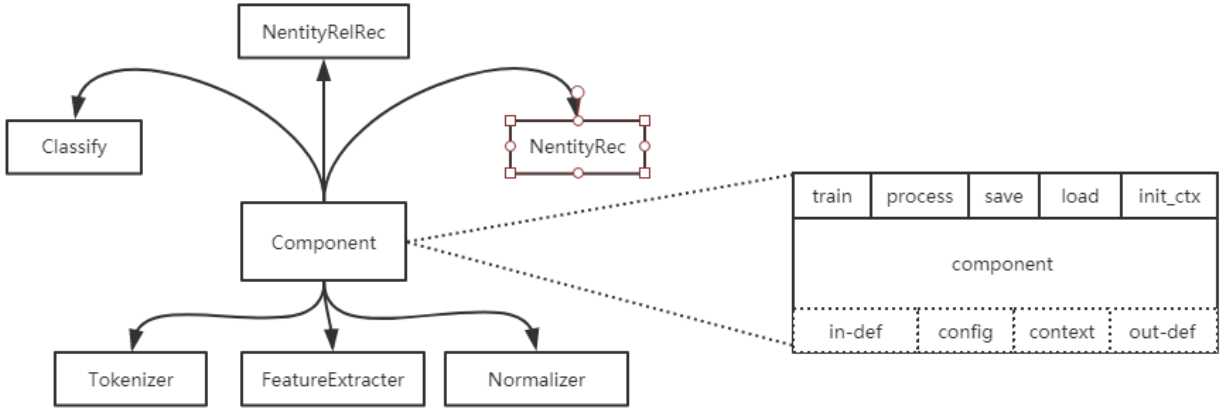
账户管理
实例管理
数据导入导出
统计可视化
标注进度
etc.

Software Engineering Matters

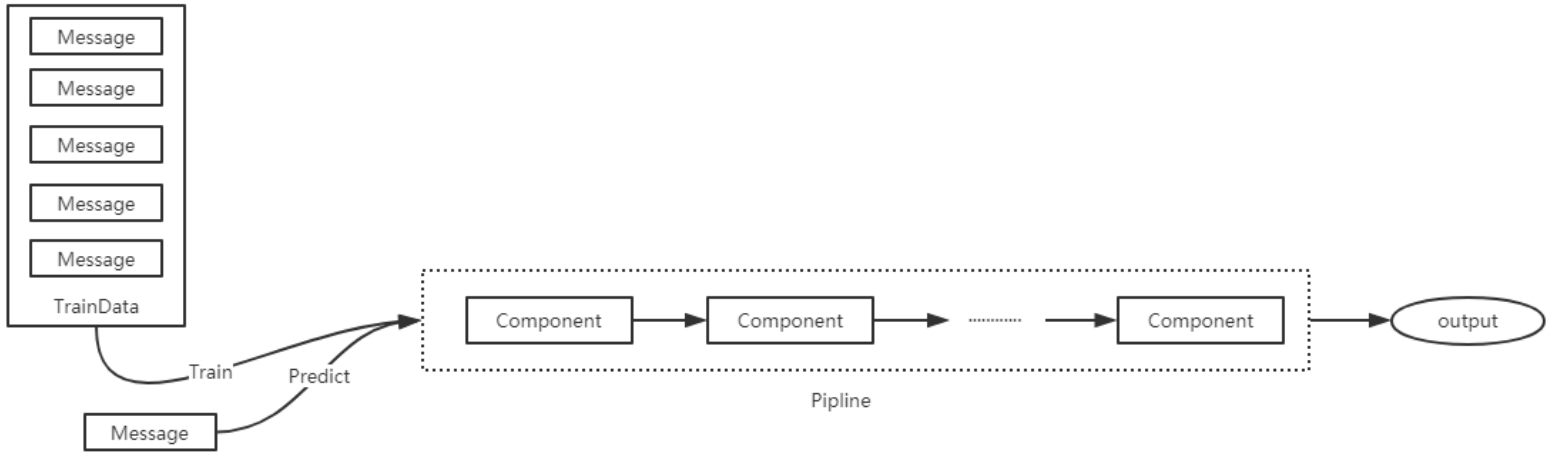


Software Engineering Matters

Modular Design



Data Pipeline Design



Software Engineering Matters

Configurable Programming

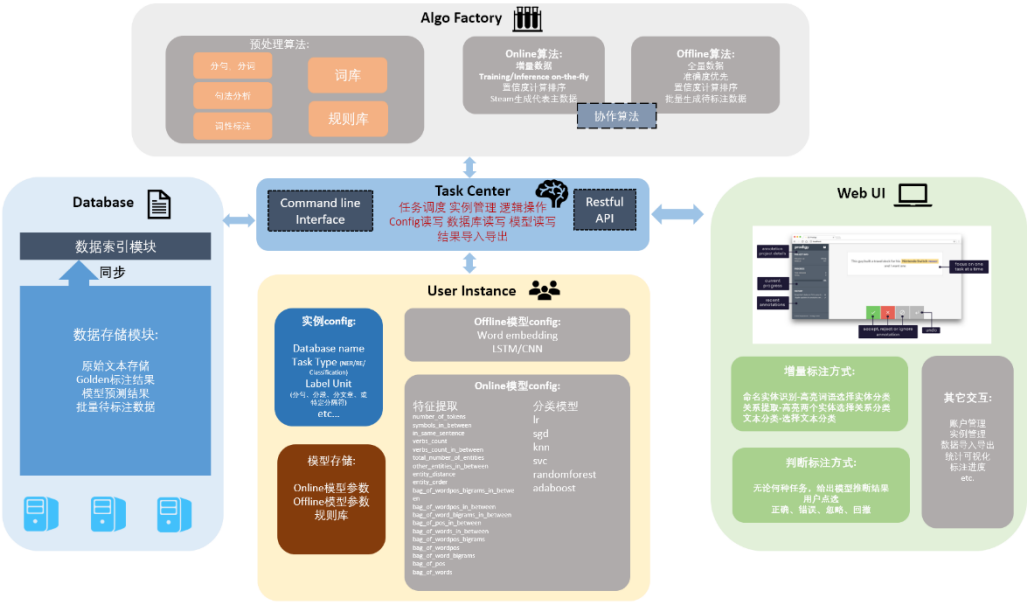
```
{
 "ip" : "localhost",
 "port": "8000",
 "database_type": "mongodb",
 "type": "classification"
 "name": "email_spam_classification",
 "model_type" : "classification",
 "pipeline": ["nlp_word2vec",
            "linesplit_preprocess",
            "feature_extractor",
            "online_svm_classifier_sklearn",
            "offline_svm_classifier_sklearn"],
 "language": "zh",
 "wordvec_file": "./tests/data/test_embedding/vec.txt",
 "path" : "./tests/models",
 "org_data" : "./tests/data/test_email_classify/email_classify_chi.txt",
 "database_name" : "spam_emails_chi",
 "labels": ["spam","notspam"],
 "batch_num" : "10",
 "inference_num" : "20",
 "low_conf_num" : "10",
 "confidence_threshold" : "0.95",
 "log_level": "INFO",
 "log_file": null
}
```

Software Engineering Matters: More than an Annotation Tool

Thanks to **Modular** and **API** design:

1. Human User Interface for Machine Learning Projects
2. Data Manager
   - raw data, pre-processed data, feature engineered data, labelled data, predicted data etc.
   - Upstream module like Crawlers
   - Downstream modules like Visualizations
3. Model Manager
   - pre-trained models, configuration of online and offline models, persisted models
4. Prediction Service

Full Pipeline Machine Learning Tool



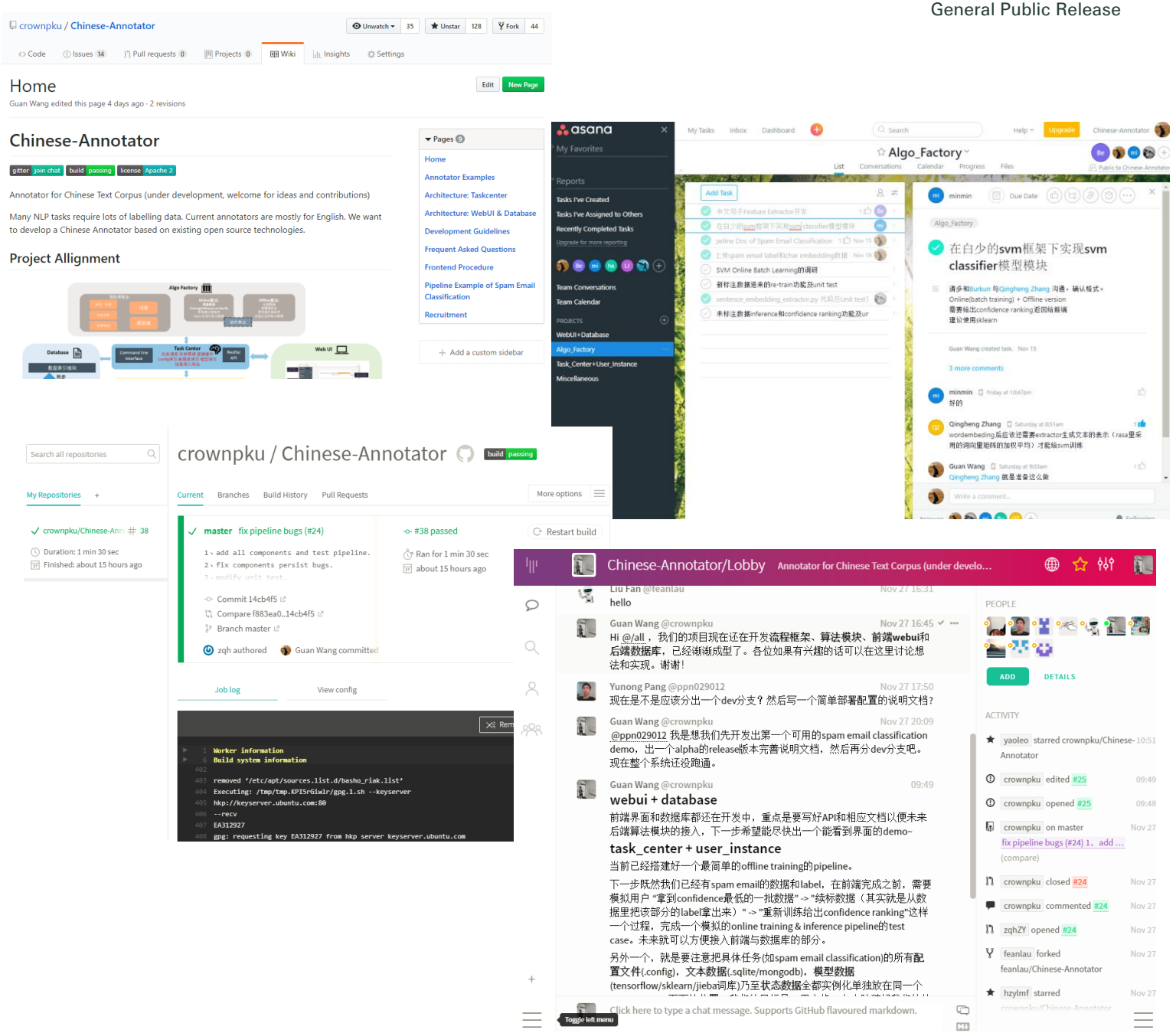
Project Management Matters

- 10 People
- Full Stack hackers, algorithm experts, professional software engineers
- Scattered in Chengdu, Shanghai, Nanjing, Beijing, Guangzhou, Shenzhen and Hong Kong (luckily same time zone)
- Process
  - Vague idea
  - White Board design
  - Software architecture diagram
  - Choice of stack back/front end
  - Configurable and Pluggable Algorithm Design
  - Unit Testing Cases
- As of July.5th 2018:
  - 30k lines of codes
  - 157 commits
  - 431 stars
  - 115 forks
  - 15 issues
  - 41 pull requests.

Project Management Matters

Tools we use:

- Github
  - Code review
  - issue tracker
  - code discussion
  - wiki documents
  - Travis-CI (continuous integration)
- Asana
  - project management
  - allocate and claim tasks
  - set deadlines
  - synchronize progress
  - share resoures
  - discuss spedific tasks
- Wechat (core team discussion)
- Gitter chatroom (public discussion)



Project Management Matters

- Working Style: **Distributed Asynchronous Collaboration**
- No compulsory meetings, every communication is online and persisted in an asynchronous way
  - Pull requests at mid-night
  - Merge and review code on the subway
- Every member chooses when, where and how he works.
  - beds, sofas, cafes, libraries, swimming pool or anywhere with good quality Wifi. No commute.
  - Save rent; Help environment protection
- Flat and Transparent, Spontaneous Collaboration, Full Trust and Support for Human Nature

Legal notice

The information and opinions contained in the presentation are provided as at the date of the presentation and are subject to change without notice. Although the information used was taken from reliable sources, Swiss Re does not accept any responsibility for the accuracy or comprehensiveness of the details given. All liability for the accuracy and completeness thereof or for any damage or loss resulting from the use of the information contained in this presentation is expressly excluded. Under no circumstances shall Swiss Re or its Group companies be liable for any financial or consequential loss relating to this presentation.