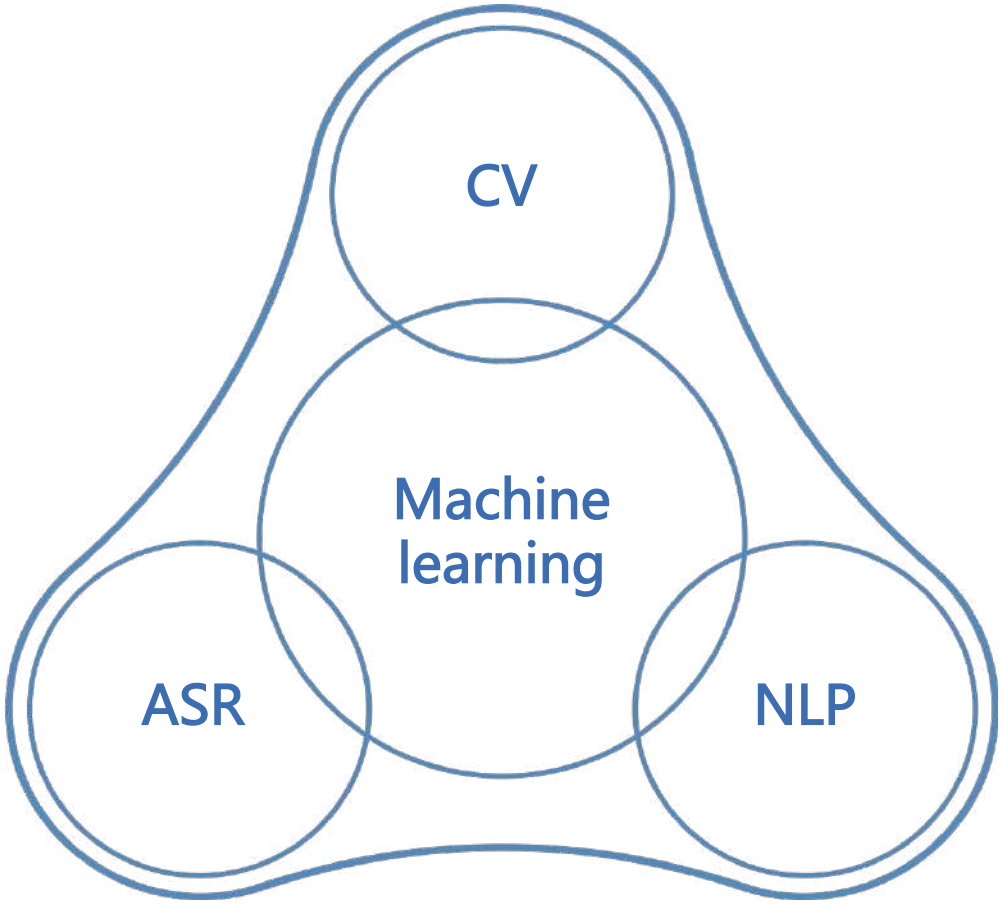
Tencent AI Lab

腾讯的自然语言应用和研究

2017 / 07

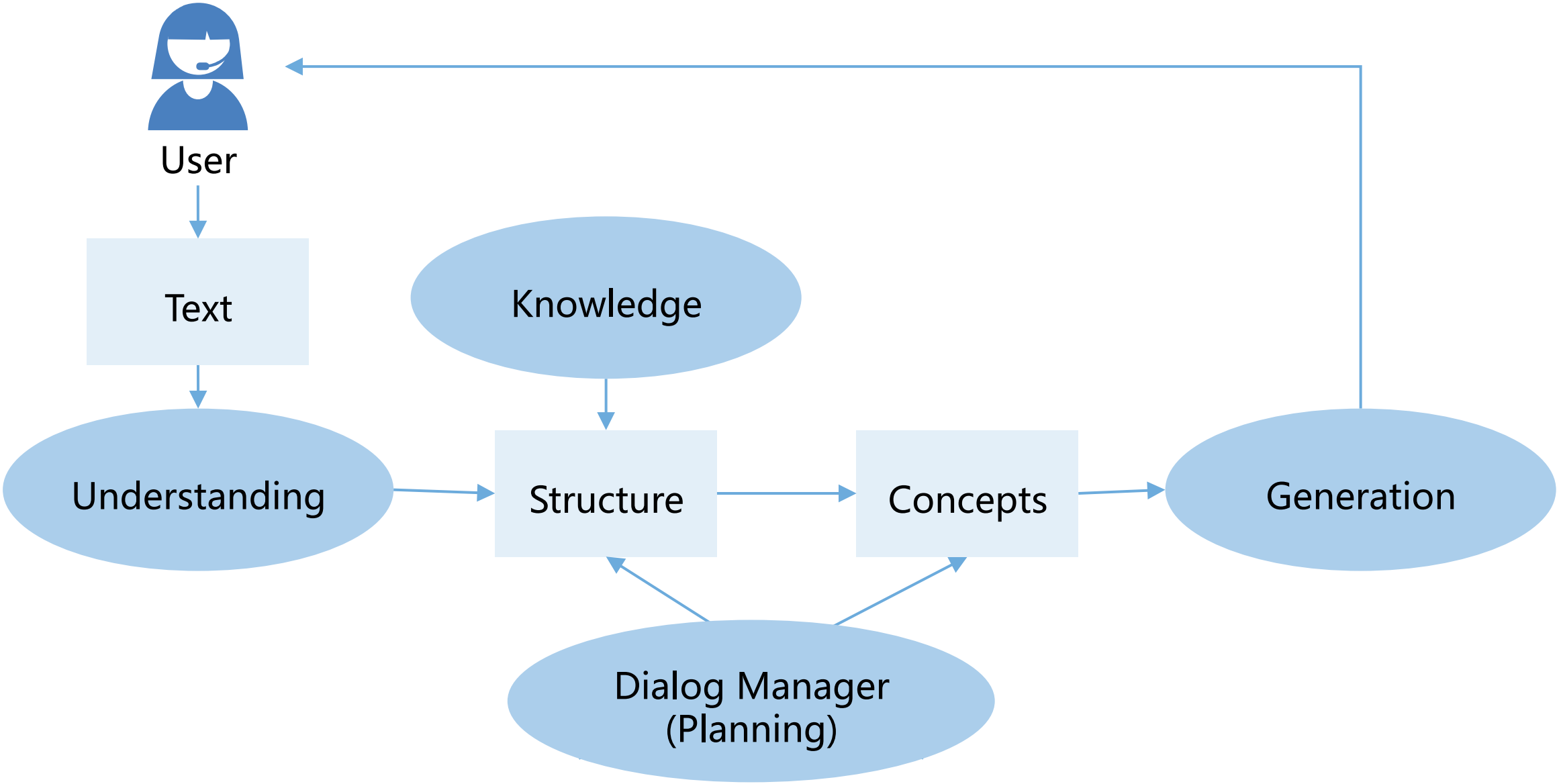
Tencent AI Lab: Fundamental Research Areas



NLP Overview

Knowledge	Understanding	Generation	Planning
• Structured • Unstructured • Real world	• Annotation • Semantics • Matching	• Concepts → Text	• Dialog interaction • Story telling

Tencent AI NLP Dialog System



NLP Examples from Tencent AI Lab

1 Knowledge	2 Understanding	3 Generation	4 Context
Knowledge graph	Sentiment analysis		
	Search		
	News recommendation		
		Machine translation	
Chatbots			

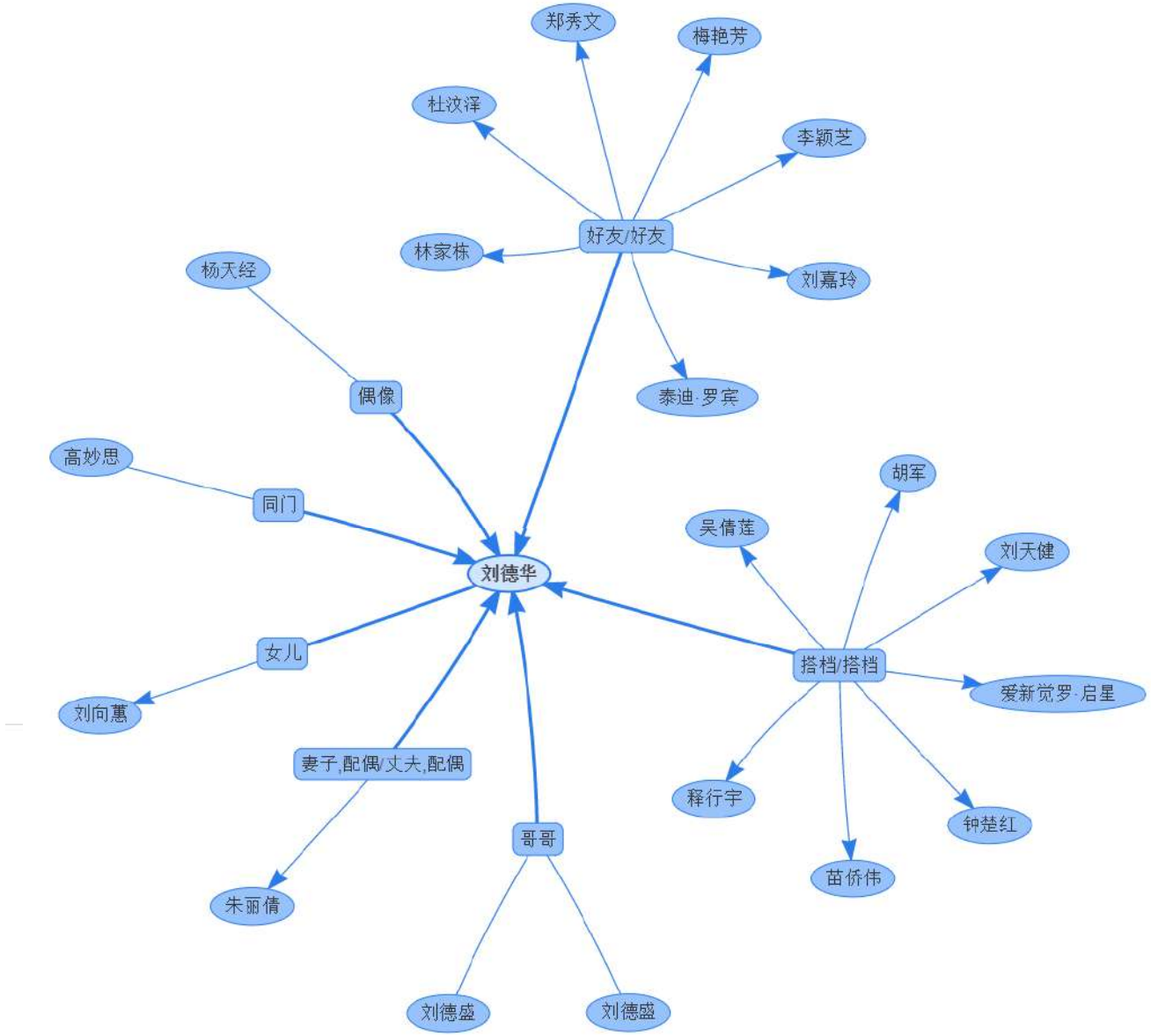
1

Tencent Knowledge Graph: billions of entities/relations

[刘德华]

导入融合覆盖集　增删改查

属性	属性值
实体名	刘德华
rich_name	刘德华
index_flag	1
实体类型	音乐类_歌手|42 视频类_影视明星|63
实体ID	11454546192129202571
popular	903
foundin(46) type_list(46)	http://baike.baidu.com/item/%E5%88%98%E5%BE%B7%E5%8D%8E 15713728408382449773 57 http://baike.baidu.com/item/%E5%88%98%E5%BE%B7%E5%8D%8E/114923 10982240448381647422 57 http://baike.baidu.com/item/%E5%8D%8E%E4%BB%94 3245896049007025704 57 http://baike.baidu.com/item/%E5%8D%8E%E4%BB%94/201395 7572876637608491092 57 http://baike.baidu.com/item/Andy



Example Usage: Q&A

- Basic Query Understanding
  - Parsing
  - Pattern-based approach
  - Attribute classification

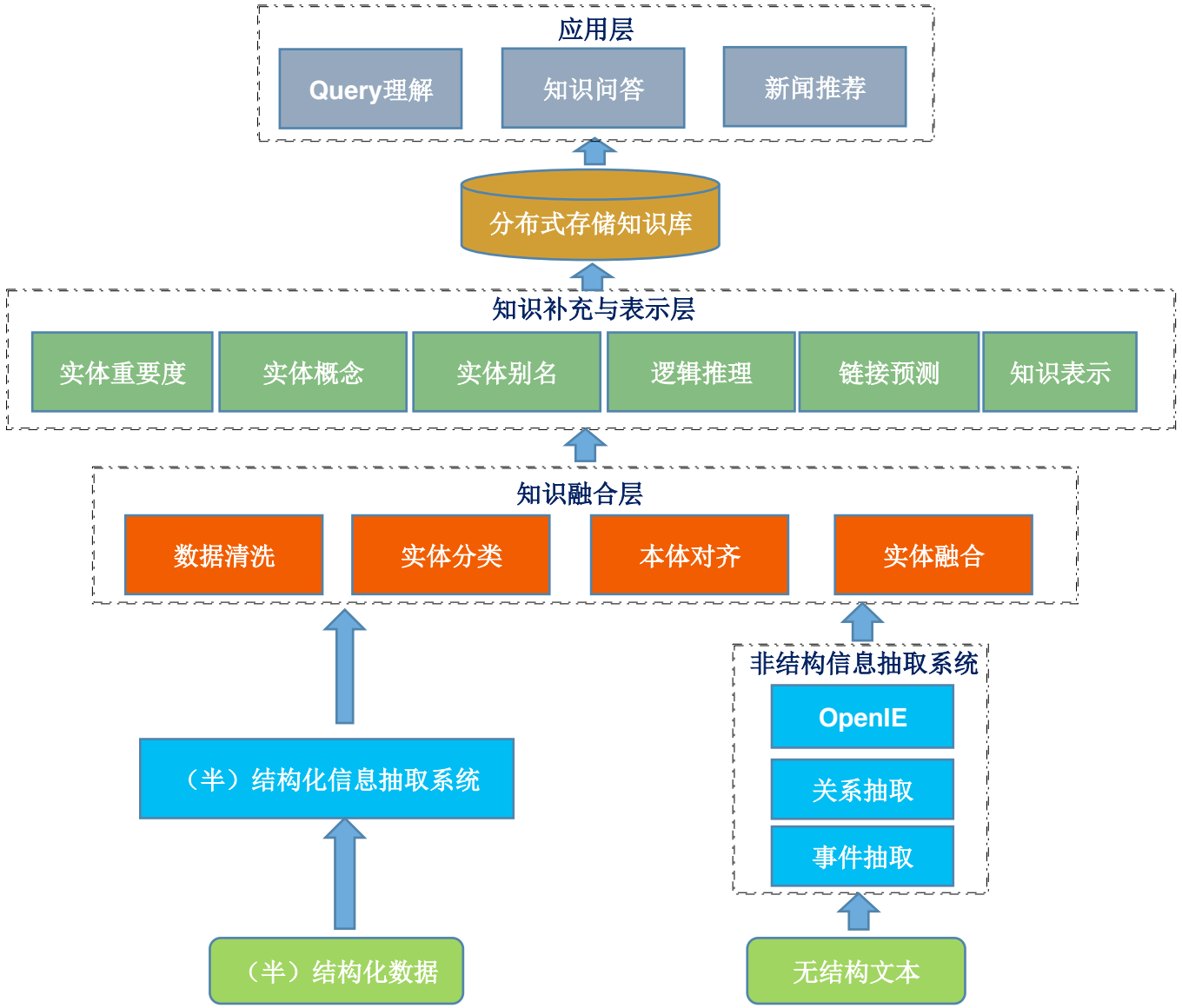
刘德华的老婆是谁
刘德华的妻子是谁
刘德华和谁结的婚
⇨ 人物类妻子属性

刘德华的出生日期
刘德华什么时候生的
刘德华几年几月生的
⇨ 人物类出生日期属性

- Complex Query Understanding
  - Complex rule engine
  - Q&A pair semantic matching

1

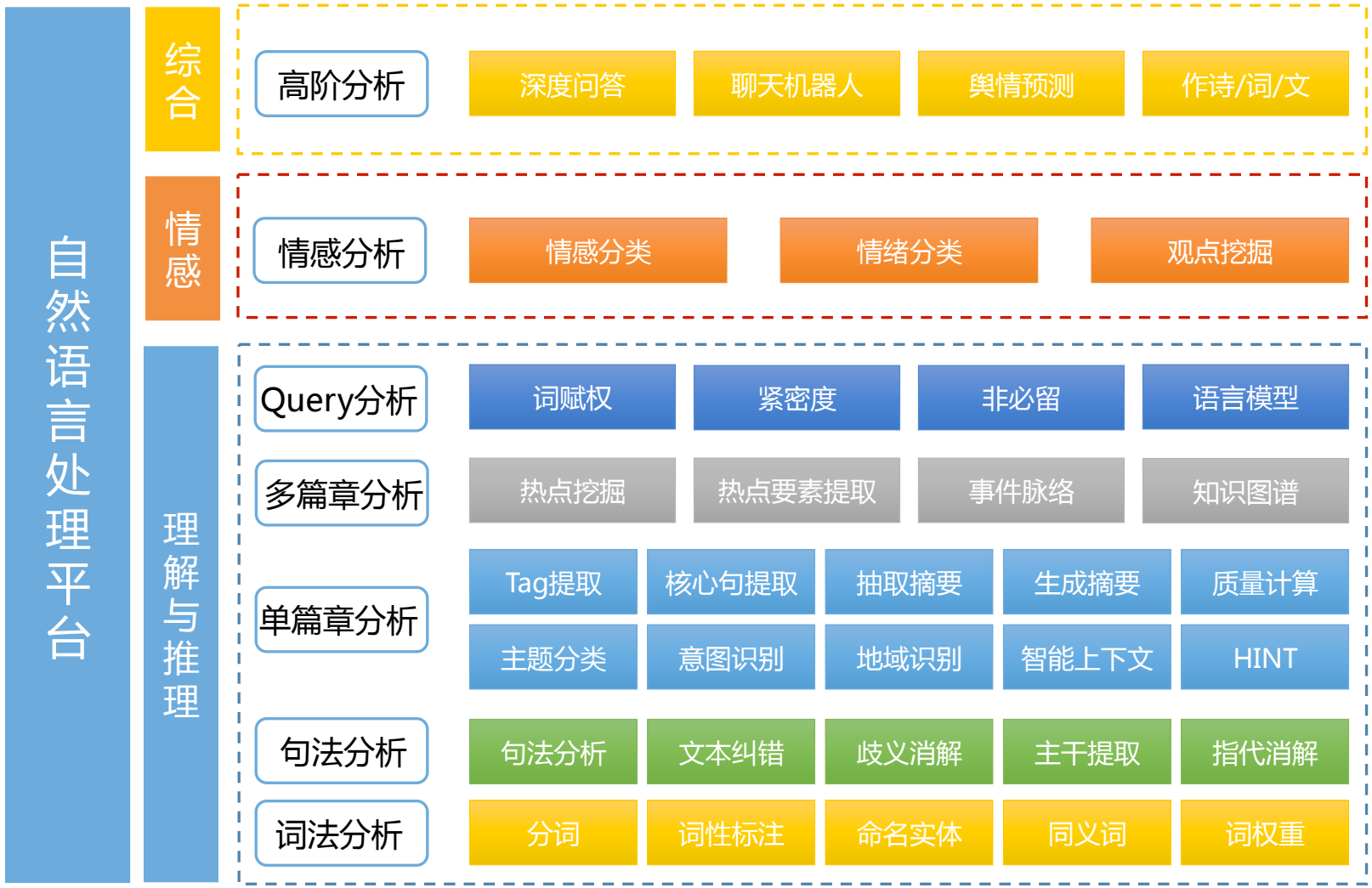
System Architecture



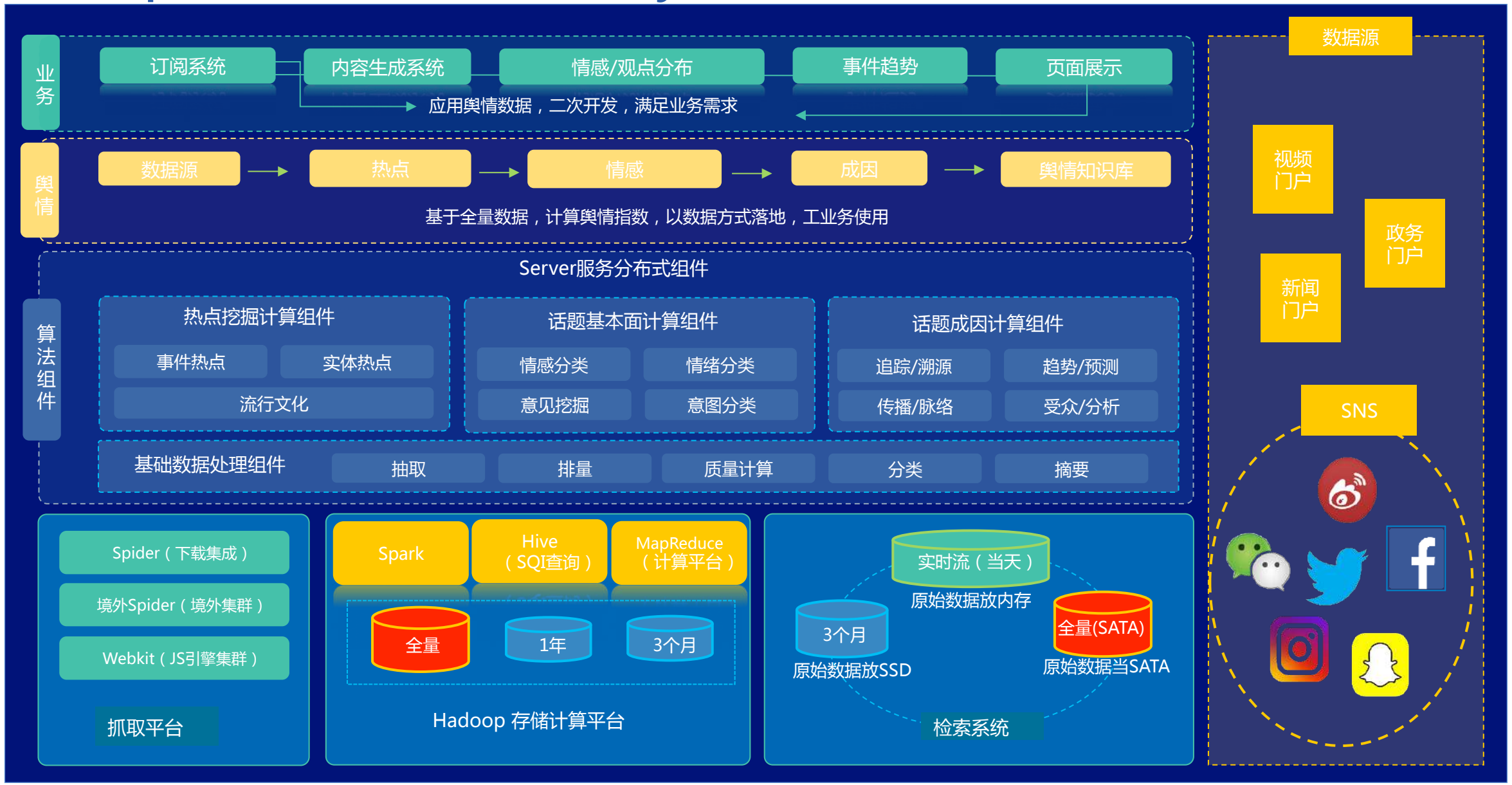
Some Key Technologies

- Entity Extraction
- Concepts Extraction
- Entity Linking
- Relation Extraction
- Ambiguity Resolution
- …

Tencent NLP Platform



Example: Sentiment Analysis Architecture



2

Use Case: People's Daily 人民日报

2

Use Case: People's Daily 人民日报

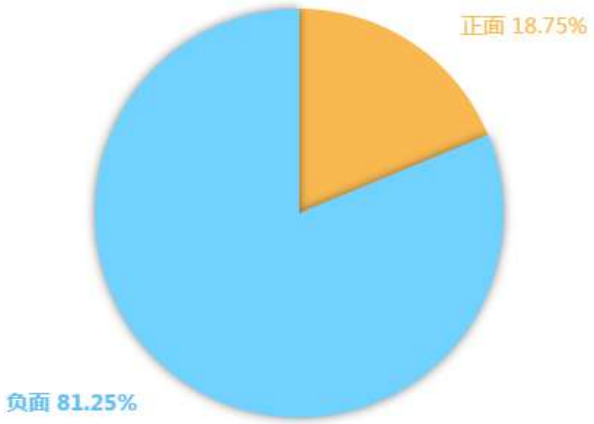
返回 深圳西乡沃尔玛发生砍人事件 分析报告

事件简介

深圳市公安局宝安分局官微@平安宝安7月16日深夜发出警情通报，2017年7月16日21时许，宝安区西乡金港华庭沃尔玛超市内，一名男子持菜刀伤人。目前已有2人死亡，9人受伤。

热度走势 情感走势 用户观点 用户画像 脉络分析 相关热词 关联素材 关系图谱

情感分布 2017-07-16 至 2017-07-17 近7天 近30天

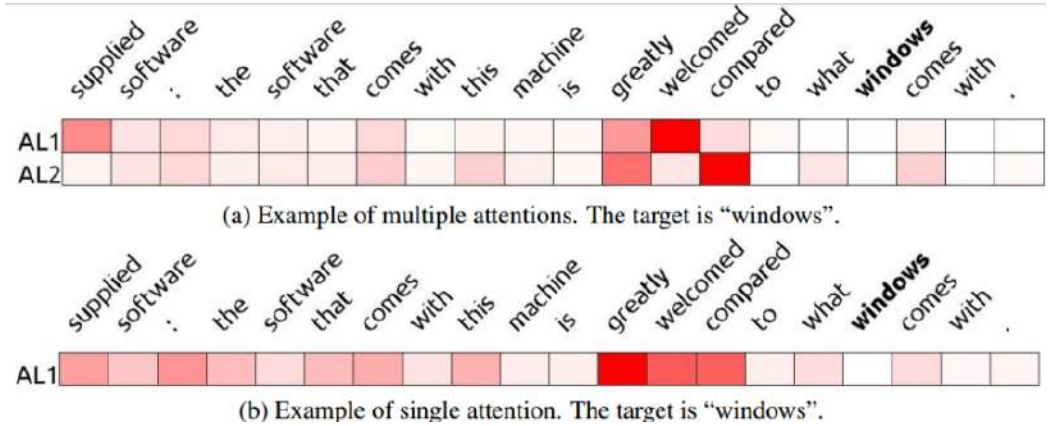


TOP 10 负面观点

排名	观点
1	不可能有是精神病吧？直接枪毙吧！太可恨了...
2	如果是心有积怨报复社会请别伤害无辜百姓，...
3	应该还有同伙吧，不可能一个人，恐怖分子真...
4	把刘死鱼砍死你就是几亿人的英雄。乱杀无辜...
5	新闻都是骗人的，有人放火有人砍人。死了4个...
6	听得都觉得很恐怖，可怜那孩子了！！她家人...
7	现在有些人就这样，心理变态，别人又没有招...
8	砍死？不是说都是受伤了吗？为什么还有砍死...
9	精神病为什么不管好？为什么没人看护？为什...
10	精神病也该枪毙，凭什么精神病就可以乱砍人...

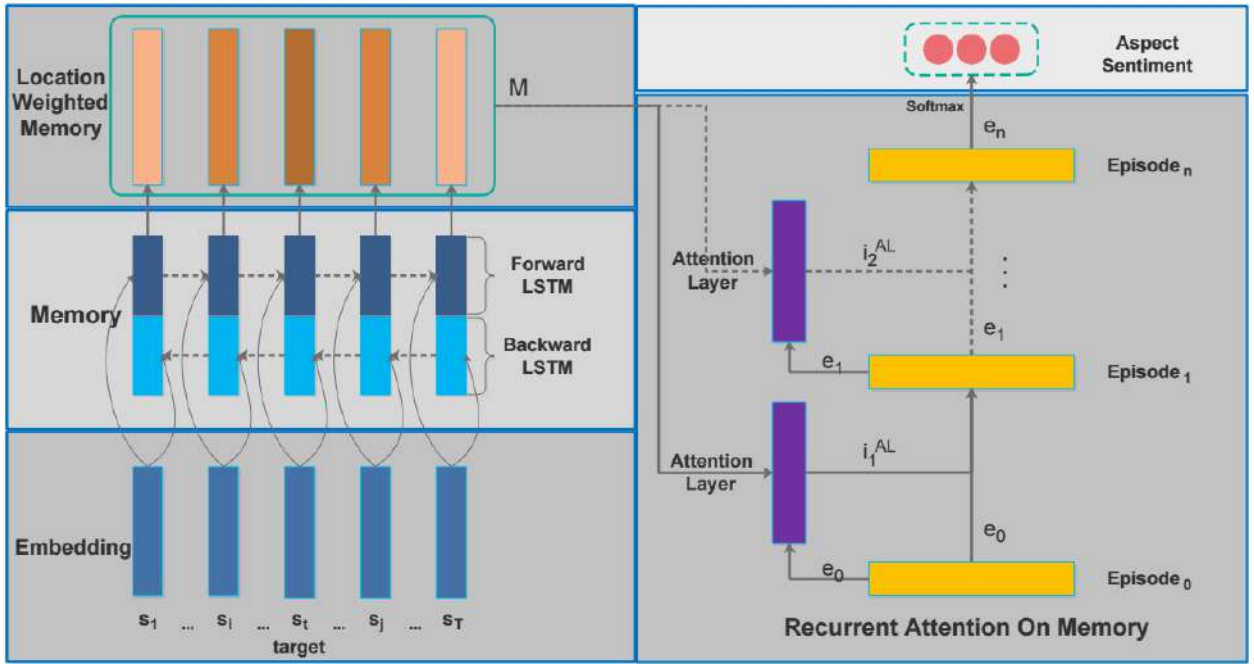
Attention Network for Aspect Sentiment Analysis [EMNLP 2017]

- Using different attentions to capture different sentiment related words
- Then, a GUR network combines these words non-linearly to predict the final sentiment



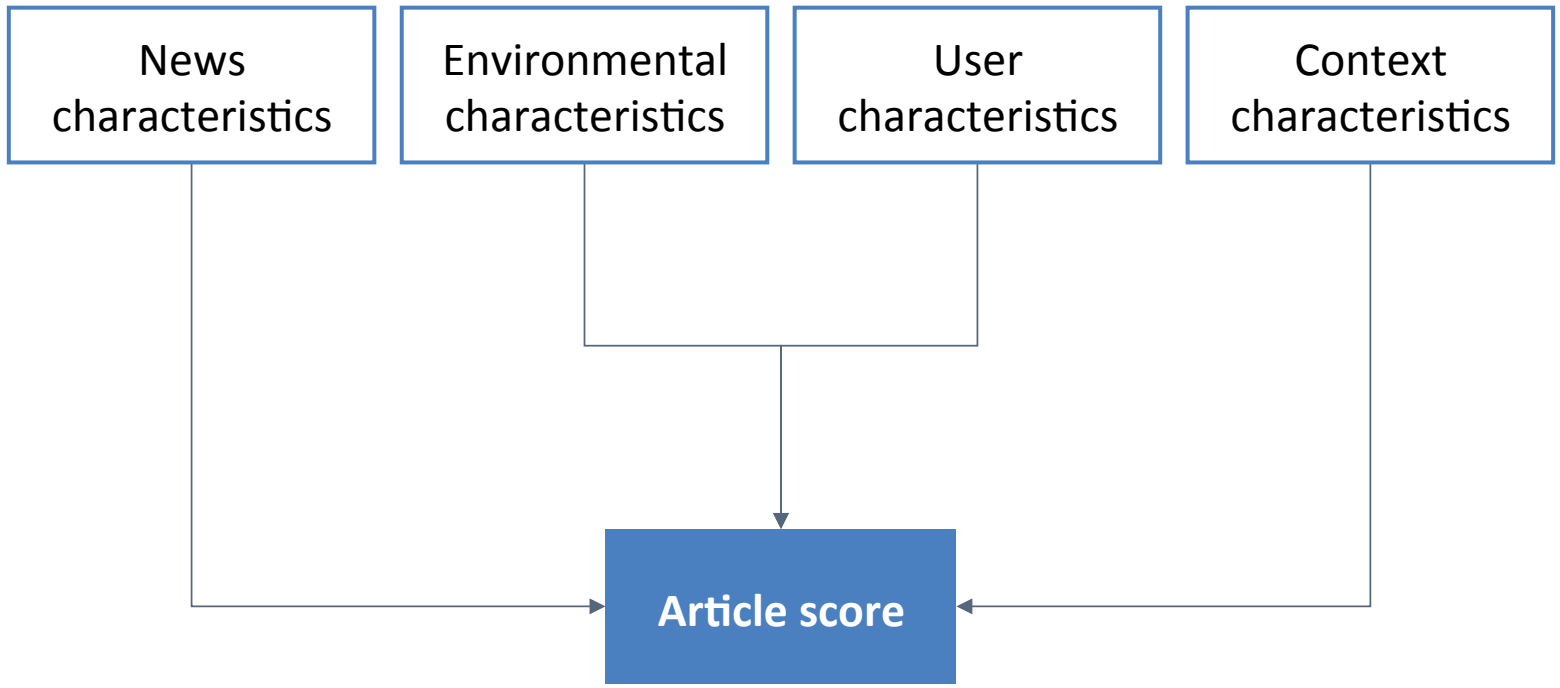
(a) Example of multiple attentions. The target is "windows".

(b) Example of single attention. The target is "windows".



Example: News Recommendation

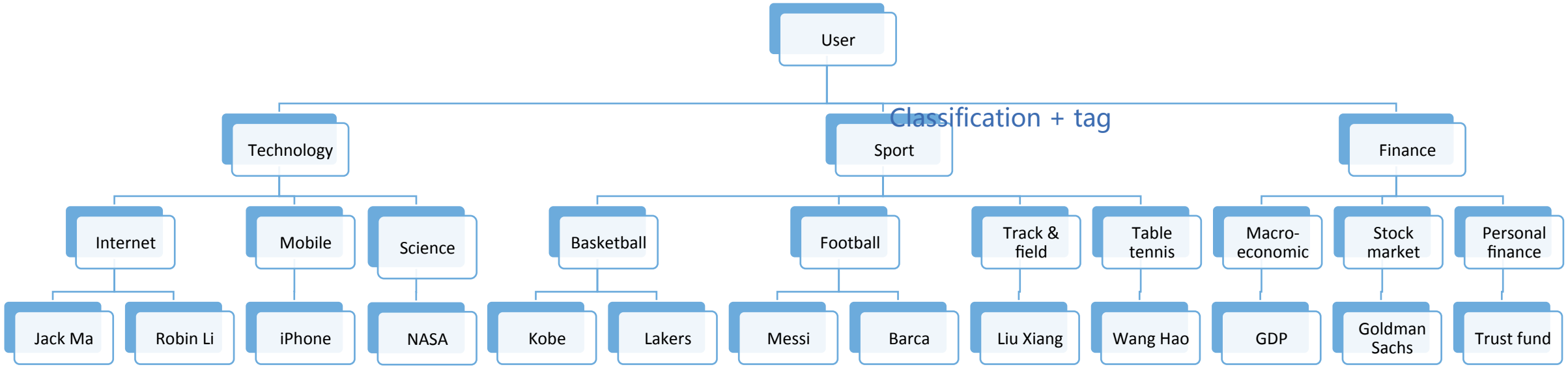
Recommendation system



Score(u,d)=f(class,topic,tag,time,…)

User Interests

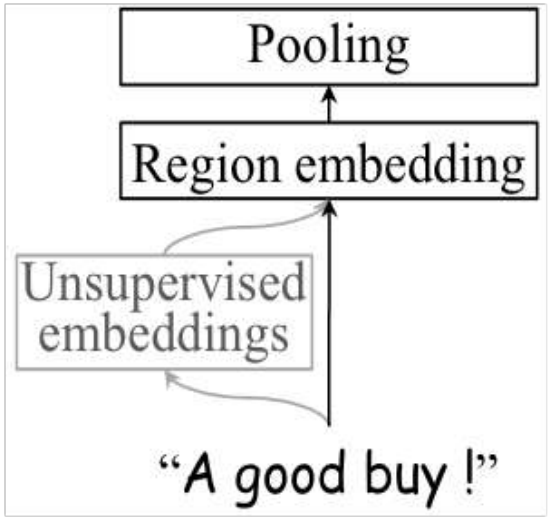
- Predefined ontology
- Automatically extracted tags
- User behavior based user interests
- …



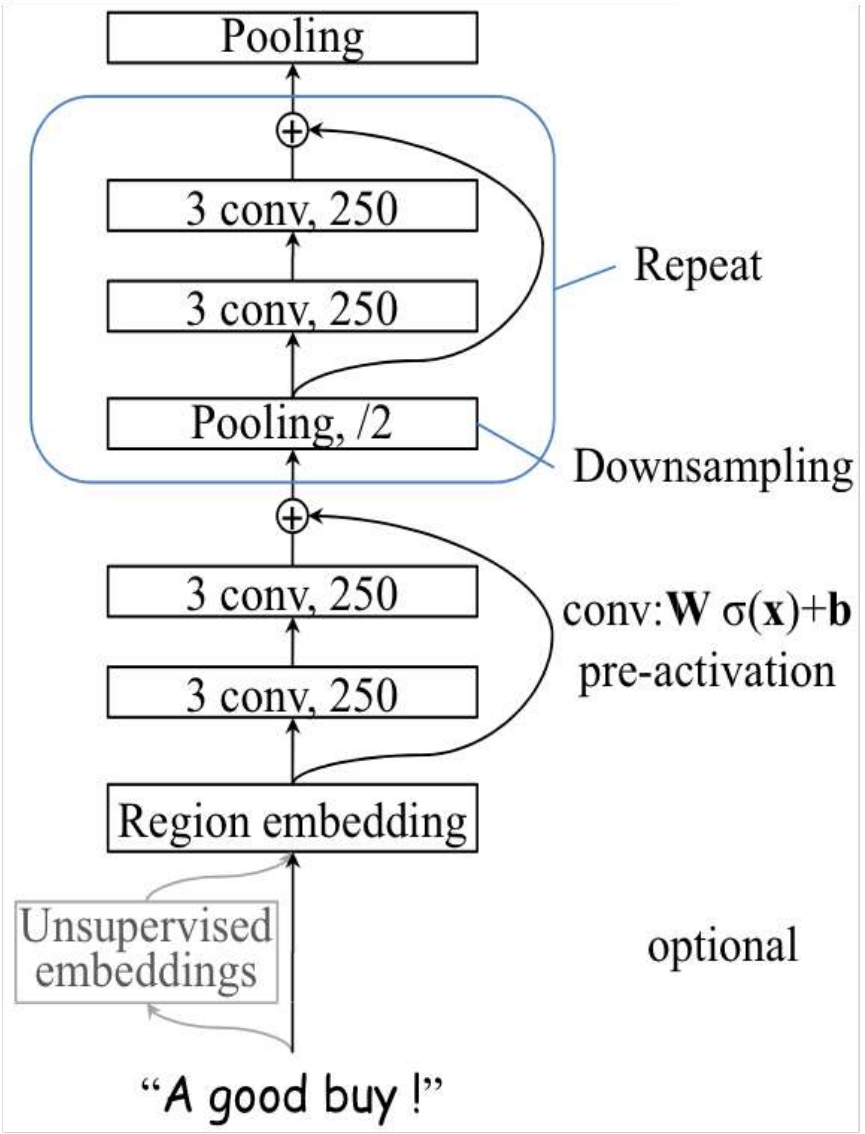
Text Categorization [ACL 17]

State-of-the-art Text Categorization Accuracy
- Deep pyramid CNN structure
- Fast computation
- Captures long range text dependency

Linear Model



Shallow CNN of (J&Z 15)



Deep Pyramid CNN (J&Z 17)

Example: Tencent Verticle Search Applications

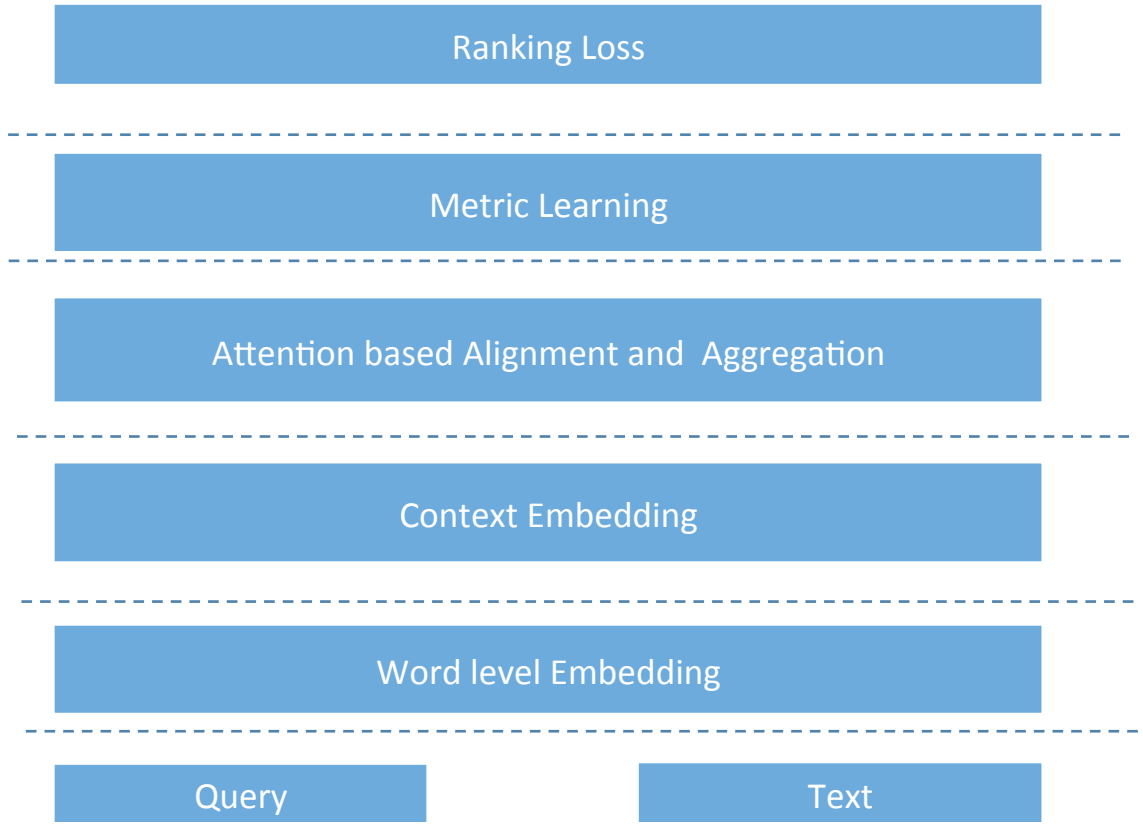
- Usage
  - E.g. WeChat search, News search, Music search, etc.
- A Major challenge
  - Quality of similarity match between queries and documents
- Technology
  - User behavior based semantic embedding

Deep Semantic Similarity Model

Example:

如何给宝宝添加辅食	六个月宝宝怎样正确增加辅食	相似度：**0.85**

System Architecture:



Generation

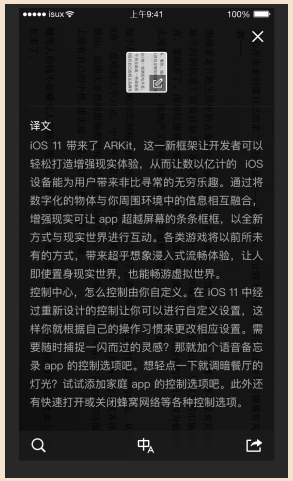
- Automatic article writing
- Summarization
- Dialogs
- Machine Translation
- …

- Key Technologies:
  - Neural Language Models
  - Memory & Attention

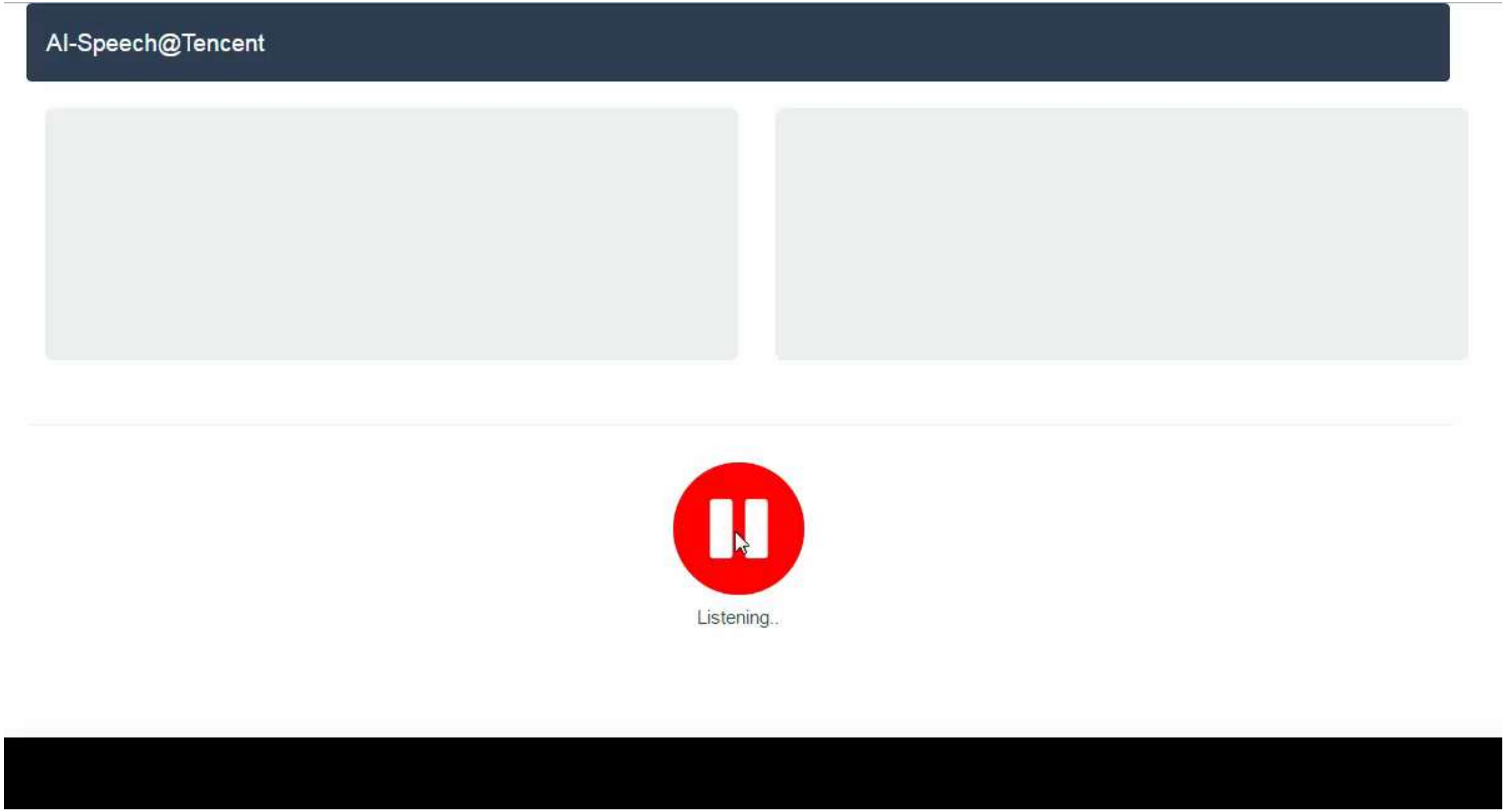
Example Machine Translation: Tencent Applications

Machine translation

Application

Text translation	+ Speech recognition Real-time conversation translation	+ Image recognition Photo translation
Translate text input	Real-time translation of speech	Translate Chinese in a photo
		
Wechat/Weixin Tencent Health Tencent Cloud AI Open Platform	APP open platform	QQ Mobile

Example: MT Video Demo



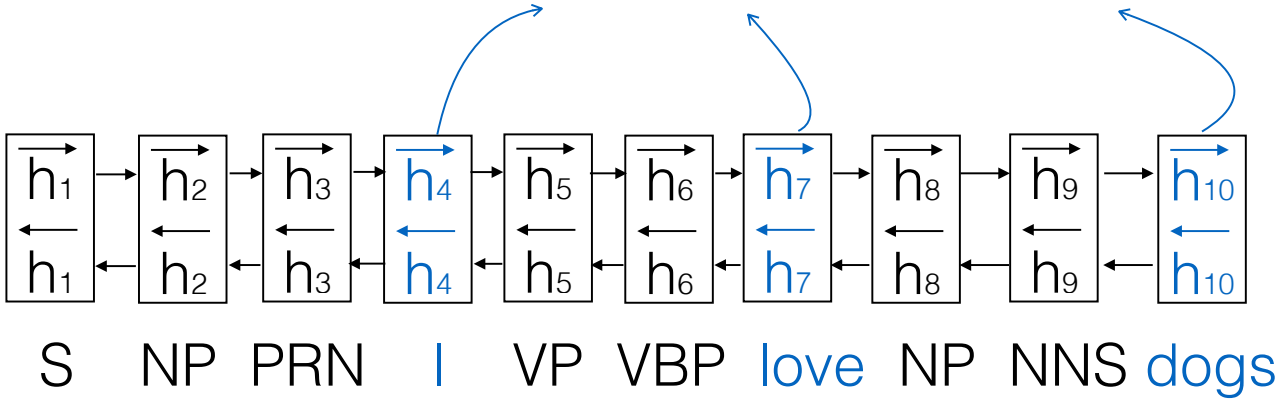
Machine Translation Challenges

- Key Challenges
  - Large Scale
  - Fluency (using NMT)
  - Under-Translation & Over-Translation

- Our Innovations:
  - Syntax-Aware
  - Multi-Granularity
  - Large Context

Modeling Source Syntax [ACL 2017]

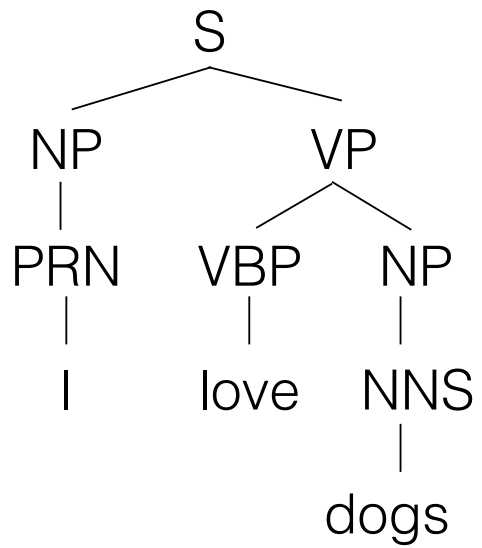
- Convert syntax tree to structural tag sequence
- Use a uniform RNN structure to generate a representation of the source language sentence containing the syntactic knowledge information



I love dogs

w1 w2 w3

(a) word sequence



(b) phrase parse tree

S NP PRN VP VBP NP NNS

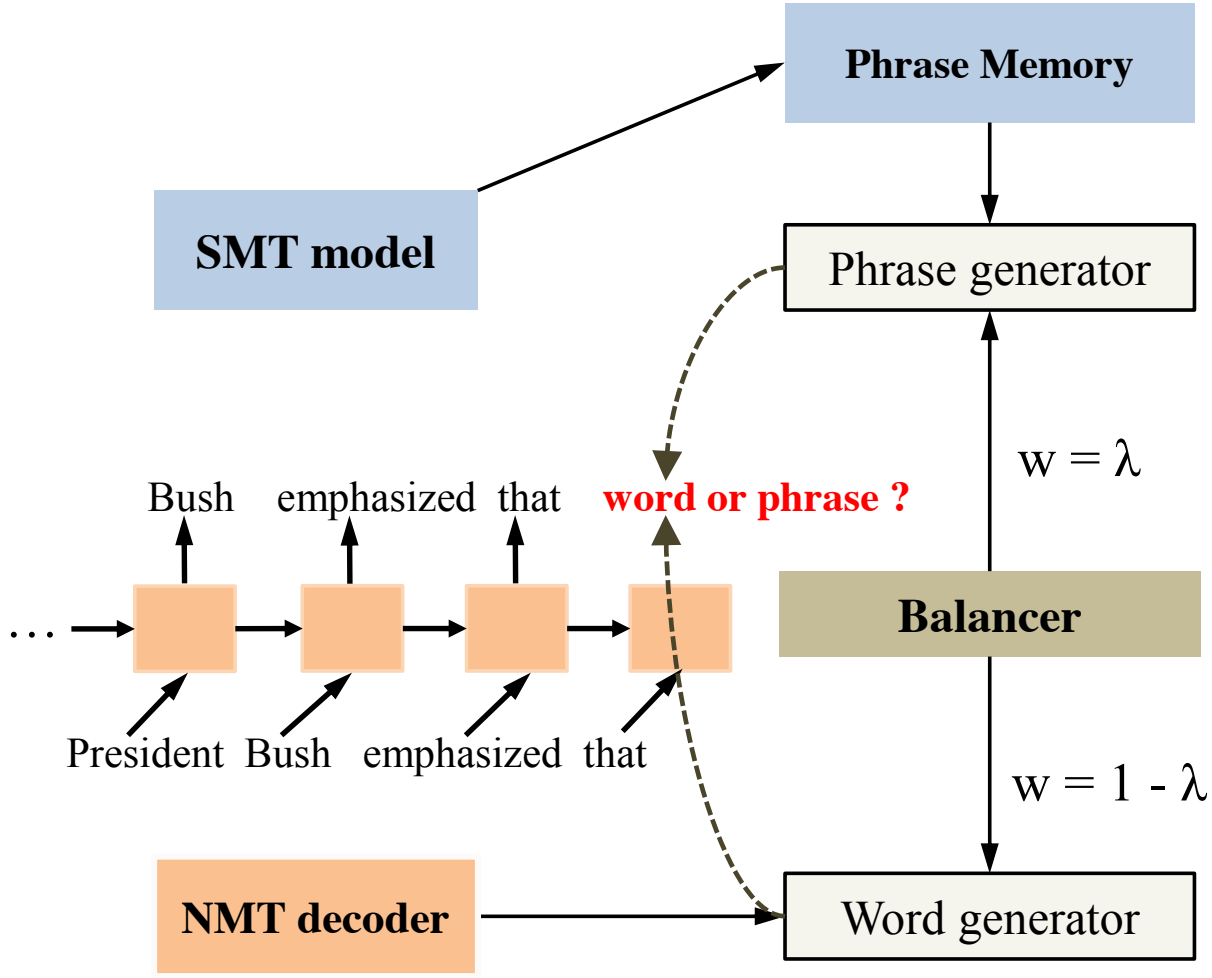
l1 l2 l3 l4 l5 l6 l7

(c) structural label sequence

Translating Phrases in NMT [EMNLP 2017a]

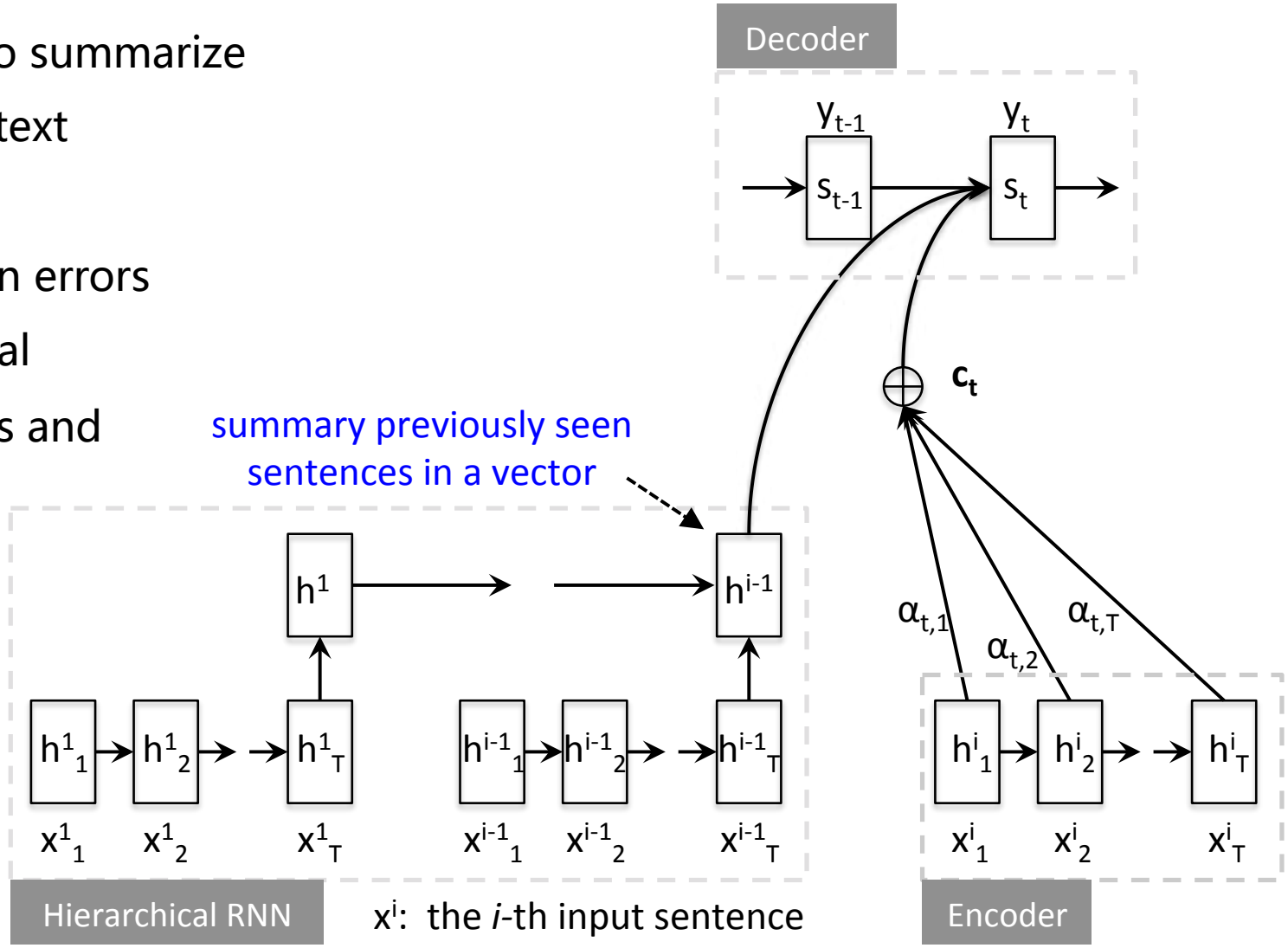
- Combine global sorting ability of neural network translation model with translation of phrases
- Improve translation ability for expressing named entities, multiple words

腾讯成立了人工智能研究部门腾讯AI Lab。
Tencent has set up **artificial intelligence research department** **Tencent AI Lab**.



Exploiting Cross-Sentence Context [EMNLP 2017b]

- Use recursive neural network to summarize previous text and improve context information
- Significantly reduces translation errors resulting from lack of contextual information such as ambiguous and inconsistent translation



Chatbots

Natural human computer interface

Many use cases in Tencent

Games: Glory of Kings, …

Social: QQ群助手

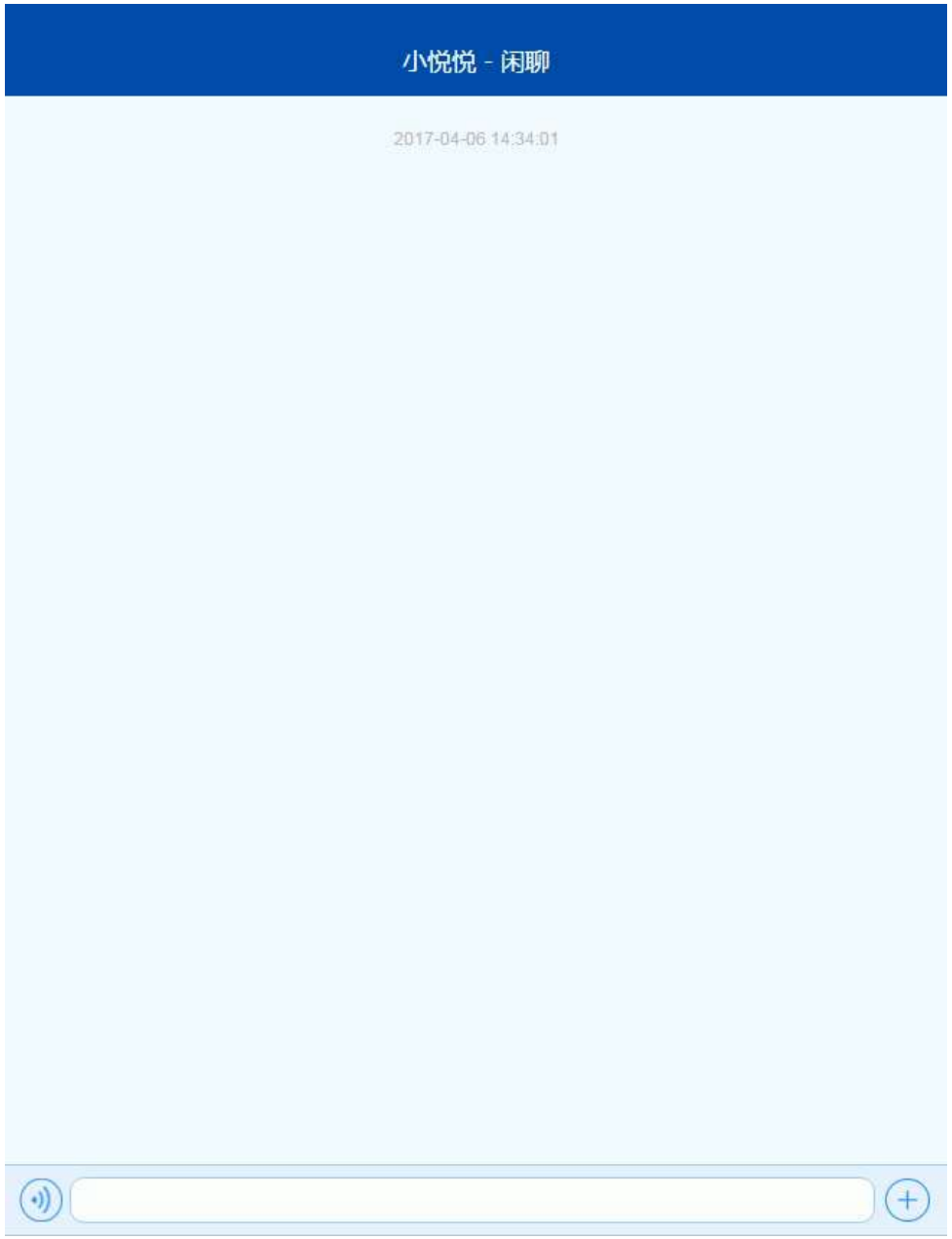
Assistants: office help desk

customer service, …

Significant Technical Challenges

Require breakthroughs in NLP

4 Example: Personalized Chatbot



Customer Service Chatbot

User

question

answer
service

Control center

Question analysis

Fundamental NLP
- Word segmentation
- Named Entity Recognition (NER)
- Dependency Parsing

Semantic analysis
- Intention recognition
- AT (Answer Type) recognition

Music

Weather

Video

.....

Question analysis

Fundamental NLP
- Component / attribute recognition
- Non-key word recognition

Context

Semantic analysis
- Relation resolution
- Command understanding
- Service status control

Merge

Meta search

Vertical search RBU

APP services

Knowledge graph

Tencent Dialog System

- Knowledge Extraction and Semantic Matching
- Context Understanding
- Variation Generation
- Structured Language Generation
- Human Guided Learning

Tencent AI Lab Research Publications in NLP

Conference	Topic
ACL 2017	Modeling Source Syntax for Neural Machine Translation
	Chunk-Based Bi-Scale Decoder for Neural Machine Translation
	Deep Pyramid Convolutional Neural Networks for Text Categorization
EMNLP 2017	Deep Neural Solver for Math Word Problems
	Learning Fine-Grained Expressions to Solve Math Word Problems
	Exploiting Cross-Sentence Context for Neural Machine Translation
	Translating Phrases in Neural Machine Translation
	Recurrent Attention Networks for Aspect Sentiment Analysis
	Abstractive Document Summarization via Hierarchical Memory Networks
	A Deep Recurrent Generative Decoder for Abstractive Text Summarization

Our Mission

Make AI Everywhere

Research

Product

Open

Tencent AI Lab Locations

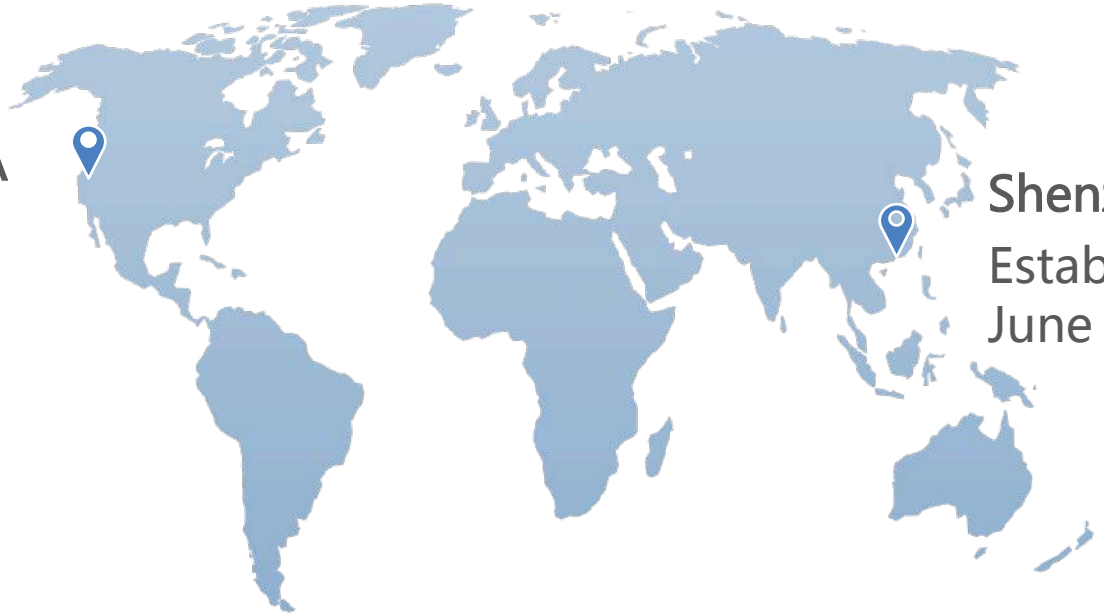
Research talent

~50 world-class scientists, researchers and experts

+

Engineering talent

~200 leading research engineers



Seattle, USA

Established
May 2017

Shenzhen, China

Established
June 2016

Thank you