

一个Google系的创业公司 如何在中国做人工智能



雷欣，CTO @出门问问 | Ticwear



我们今天聊什么？

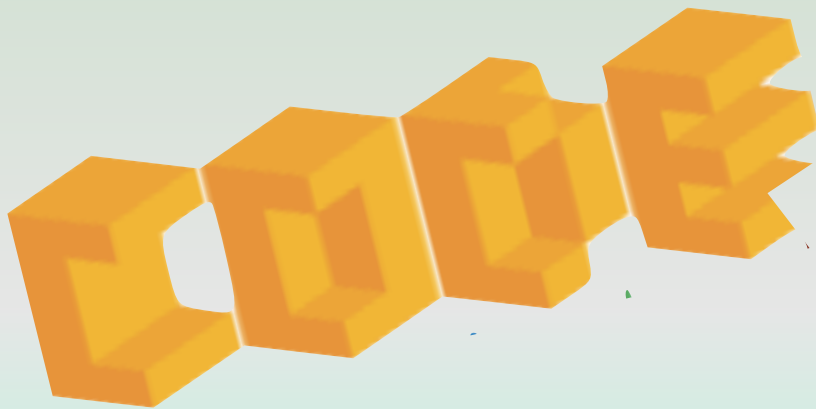
- Google的工程师文化
- 打造一支Google工程师文化的团队
- 人工智能技术的研发
- 核心技术的落地
- 本地化服务与生态系统



Google的工程师文化

工程师文化 & Google Code?

- C++ Coding Style
- Readability
- Unit Test
- Code Review
- “LGTM” (Look Good To Me)



高效的开发环境

- 强大稳定的base库
- 代码构建工具blaze
 - v.s. makefile, 集中精力在代码上, 快速并行编译
 - Facebook: buck, Twitter: pants, 国内: blade
- Protocol Buffer
- Stubby RPC
- GTest, GFlags, GLog
- MapReduce, Bigtable, etc...



敏捷的开发团队

- Google有影响力的项目都是2个人的小团队开发出来的
如搜索，广告，Gmail
- “3-5个人做不好的事情，30-50人做的更烂“
- 快速开发，快速上线，快速迭代
- 一切自动化，数据驱动



如何达到Google代码质量的一点建议

坚持原则

unit test?

code review?

拥抱开源

鼓励更新
快速迭代



打造一支Google工程师文化的团队

招什么样的人

- 通才(generalist), 而不是专才(specialist)
- Smart: 分析解决问题的能力
- Get things done
- Communication

面试流程

- Phone screen + 3轮 onsite coding interview
- Whiteboard coding: 设计、算法、编码

培训

- Coding style guide, mentor, starter project
- Codelabs

活动

问问讲堂，技术分享
hackathon

Review

Self review
Peer review

在中国招人留人的困难

- Risk taking
- Temptation

快速上线，持续迭代（天下武功，唯快不破）

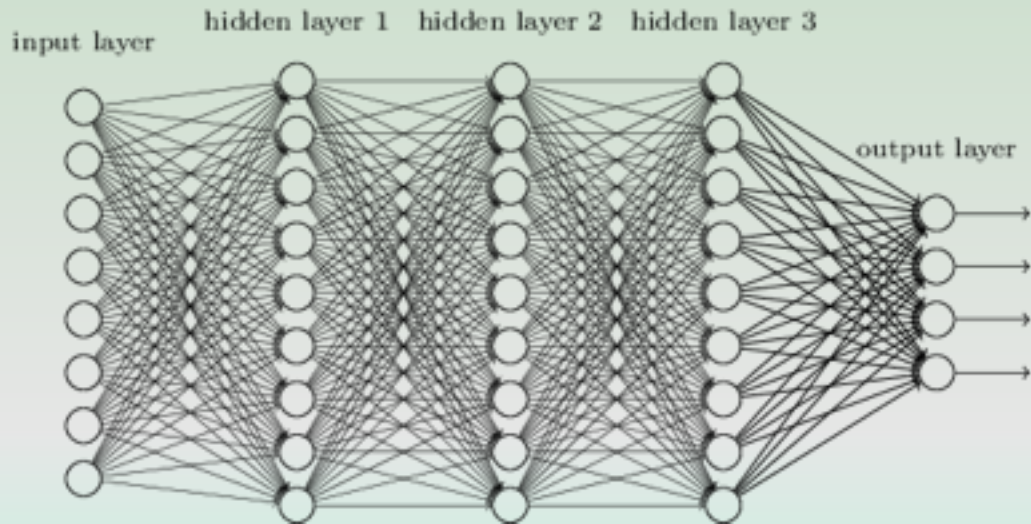
- 功能小团队：2-3人
- 每周迭代
- 提出bugs + features；周一到周三开发；周四内部测试然后粉丝内测，周五公测



人工智能技术的研发

独立研发的人工智能技术

- 语音识别
- 语义分析
- 垂直搜索
- 智能推送



“附近的**川菜馆**，**人均50块钱**左右的”

语音识别



- 基于开源，二次开发并优化
 - 低能耗热词唤醒
 - 离线语音识别
 - 云端大词汇量语音识别
- 2013年7月上线深度神经网络(DNN)系统
- 与Johns Hopkins Univ.语音研究组保持合作关系

* 语音识别研发的难点

- 训练数据的累计从零开始
- 技术上的难点
 - 噪音，远场，方言
- 商业化方向
 - 智能车载，智能家居，智能硬件
- 垂直领域深度优化：目前识别准确率90%-95%



语义分析

- 灵活的口语输入
 - “**明天下午三点左右到西安的航班，我要国航的**”
- 语义分析难点
- 规则 + 机器学习 (DNN, MaxEnt, CRF...)

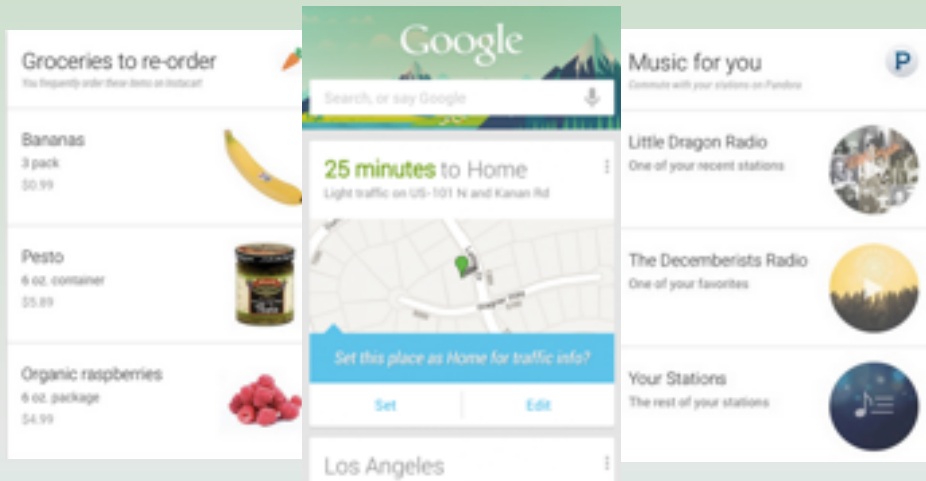
垂直搜索



- 第三方搜索API存在的问题
- 自主搭建基于网页大搜索技术的垂直搜索引擎
 - 爬虫，索引，实时检索
 - 十亿级别检索规模，秒级更新

智能推送

- 主动搜索 (Proactive Search)
- Google Now
- 常用场景
 - 下班回家：家在哪儿，实时路况
 - 火车票，飞机票提前提醒
 - 到一个陌生城市：推荐餐馆，活动
 - 新闻推荐





人工智能技术的落地

核心技术如何落地



- 出门问问微信公众账号



- 安卓手机客户端，苹果手机客户端



- Google Glass上首款中文语音搜索应用



- Android Wear上首款中文语音搜索应用

- 用户痛点？



人工智能技术的最佳载体

全球首款中文智能手表操作系统Ticwear

全中文
自然语言交互



创新Cubic UI

强大本地化服务

语

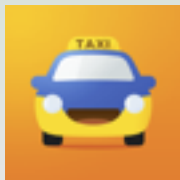
在Ticwear操作系统下 您可以语音打电话 回短信 甚至还可以回微信 打车
You can make a call, reply text message and WeChat message, even book a taxi



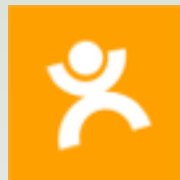
本地化服务与生态系统

本地化数据

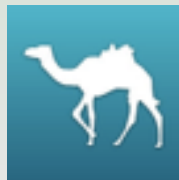
- 对接近百家本地第三方数据
- Ticwear与数十家国内顶级APP达成官方合作



滴滴打车



大众点评



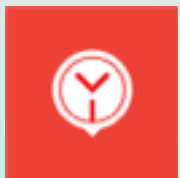
去哪儿



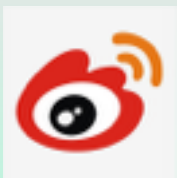
高德地图



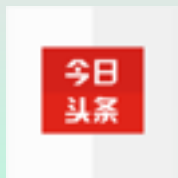
节操精选



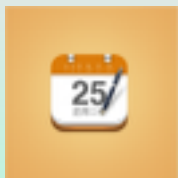
易到用车



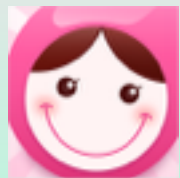
新浪微博



今日头条



中华万年历



大姨吗

打造智能手表生态系统



- 长期平台
- 面向广大开发者
- 首次开放语音语义识别及搜索API
- 定期举办线上线下黑客马拉松活动



Thank you!