大数据语义分析与应用实践

Big Data Semantic Analysis and Application

CDA 数据分析师 www.cda.cn

张华平 博士 副教授

NLPIR 大数据搜索与挖掘实验室

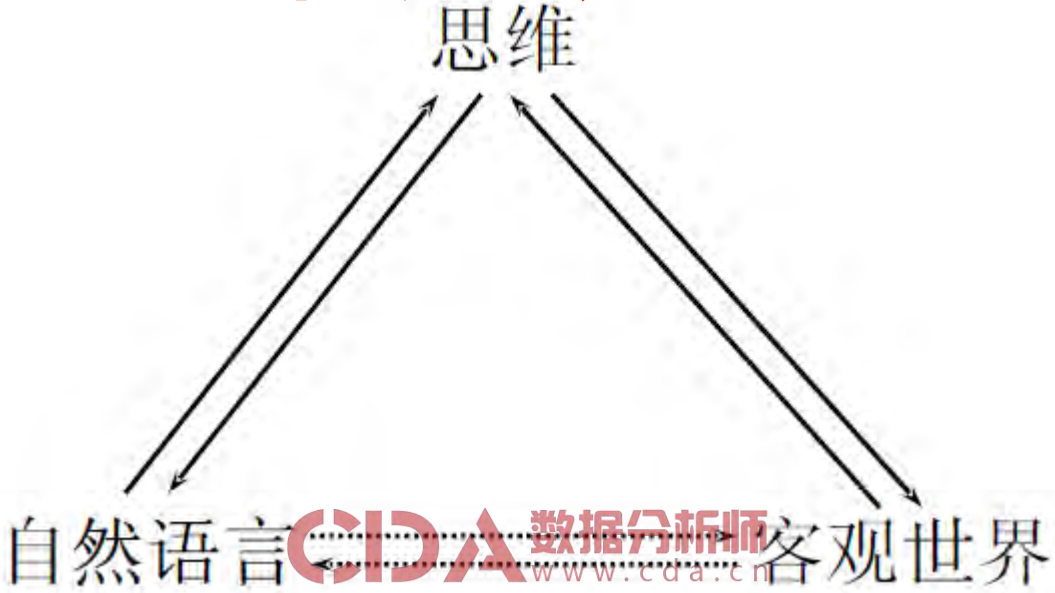
kevinzhang@bit.edu.cn

www.nlpir.org

2016.8

机器理解自然语言？

客观世界->思维->自然语言



- 衰减效应：
  - 思维最多只能反映80%的客观世界；
  - 自然语言只能反映80%的思维：词不达意，答非所问；
  - 听众最多只能听懂80%；
  - 听懂的部分只有80%能反映到思维中；
  - 分析客观世界的最多只能利用80%。

纲要

大数据搜索挖掘

I 科学的大数据观

II 文本大数据挖掘关键技术

III 大数据精准搜索关键技术

IV 大数据语义应用实践

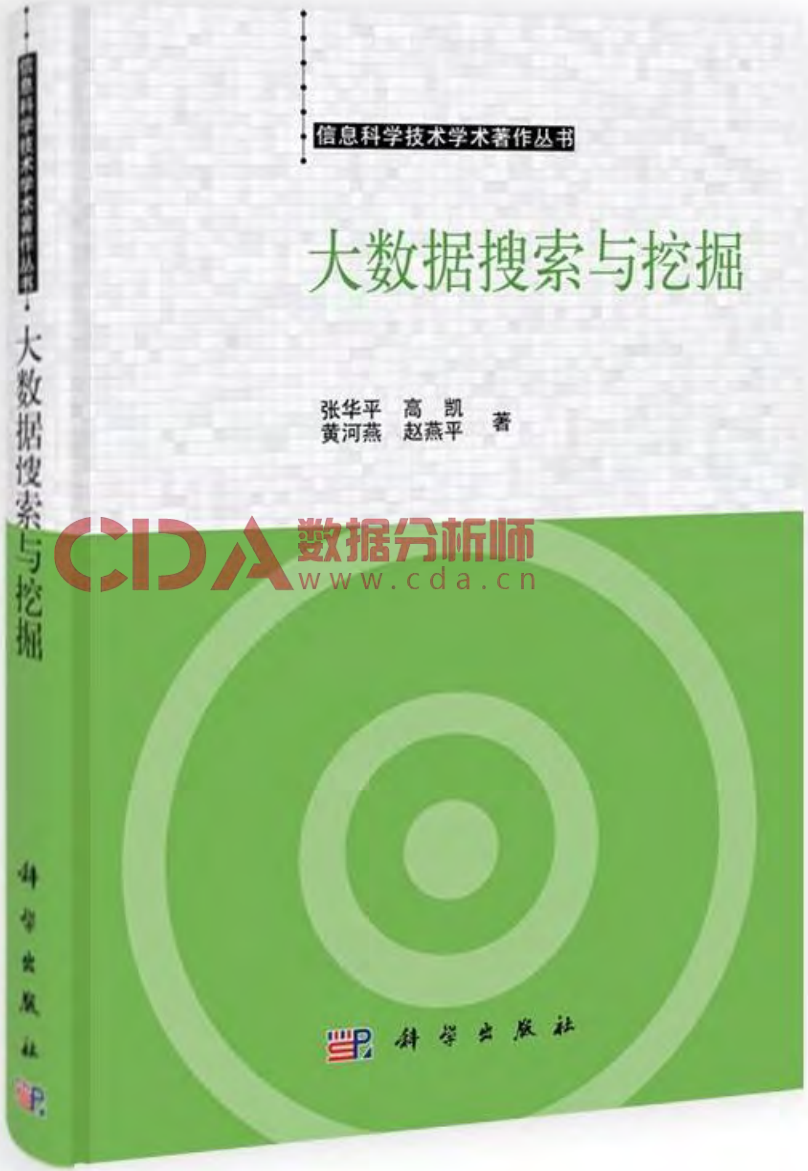
什么是大数据

- **Wiki: Big data** is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.
- 维克托《大数据时代》：大数据指不用随机分析法（抽样调查）这样的捷径，而采用所有数据的方法。

什么是大数据

我们的见解

- 大数据是指 、多源异构、实时 然语言处理、信息 知识，转化为智慧
- 是一场新的 革命。（全量分析 地认识世界的混杂 断）



杨达才启示：1+1>>2才是大数据

近半世纪来的三次革命

计算机时代

20世纪
70年代

计算方式的革命

互联网时代

20世纪
90年代

信息传播方式的革命

大数据时代

21世纪10
年代

Big Data

决策方式的革命

大数据颠覆决策模式

ion

ime

大数据时代的特征

纲要

大数据搜索挖掘

I 科学的大数据观

II 文本大数据挖掘关键技术

III 大数据精准搜索关键技术

IV 大数据语义应用实践

大数据应对之道:知著、见微、晓意

小

小 小 小 小 小 小 小 小 小

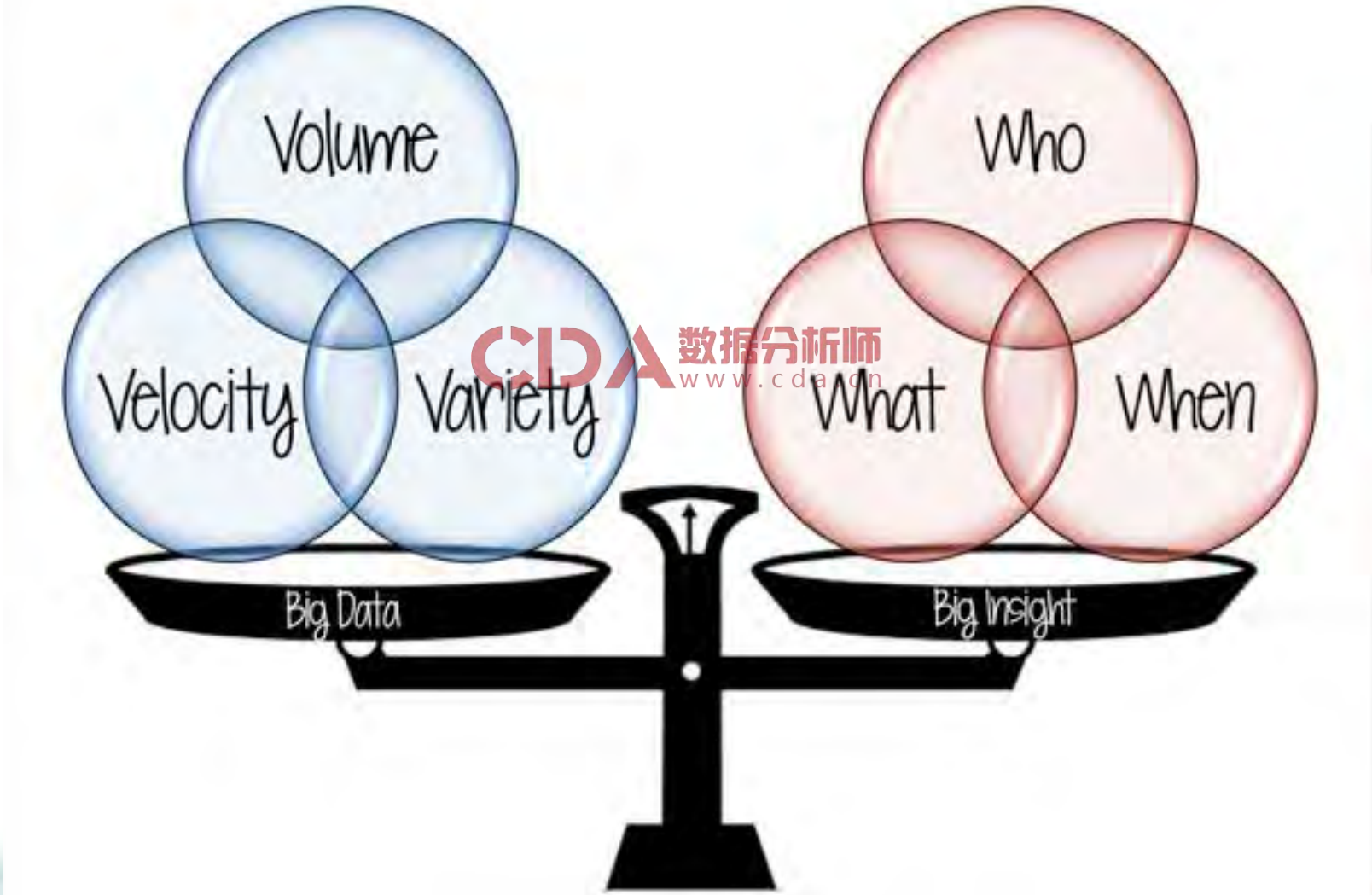
知著 小 见微

小 小

小 小

小 小 晓意 小 小

大数据更大意义上是非结构化内容理解



大数据搜索与挖掘关键技术

NLPIR大数据搜索与挖掘技术开发平台

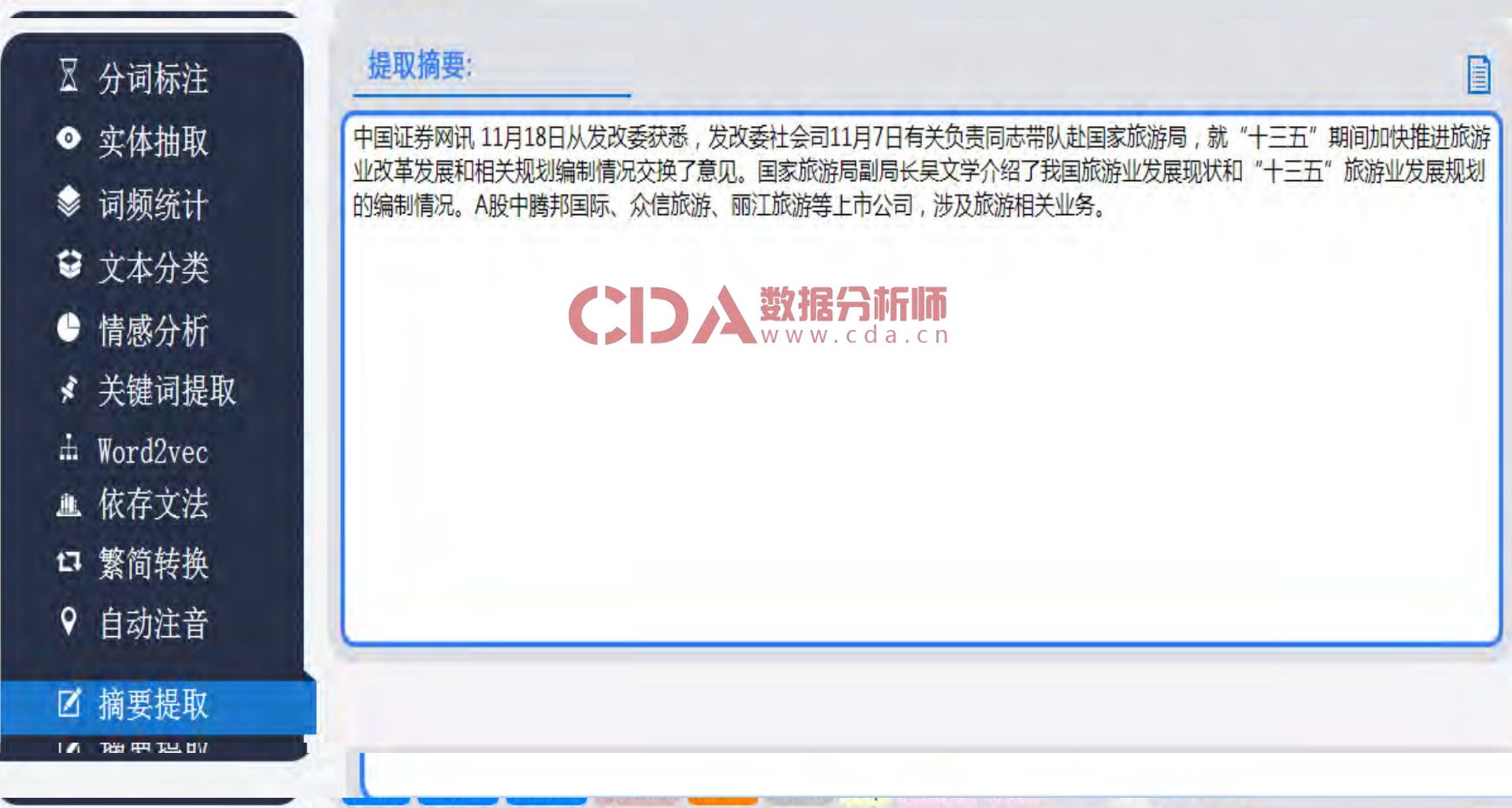
- NLPIR网络搜索与挖掘共享开发平台，针对语言信息内容处理的全技术链条的共享开发平台。15年专业研究与工程积累，提供应用软件及各平台下的二次开发包，非商用永久免费。www.nlpir.org下载。

自然语言处理与信息检索共享平台
Natural Language Processing & Information Retrieval Sharing Platform

- 核心功能包括：
  - 搜索类：全文精准检索；
  - 语言类：新词发现，分词标注，统计分析与术语翻译；关键词提取；
  - 文档类：文本聚类及热点分析；分类过滤；自动摘要；文档去重；情感分析

NLPIR大数据语义分析技术的在线演示

网址：http://ictclas.nlpir.org/nlpir/



产品下载试用

网址：https://github.com/NLPIR-team/NLPIR

..		
Classify	update Linux 64 bit NLPIR and JZSearch (CentOS)	3 months ago
Cluster	update sentiment and SentimentAnalysis	2 months ago
DeepClassifier	add IOS support, can be used in Macbook and iPhon3	14 days ago
DocExtractor	add IOS support, can be used in Macbook and iPhon3	14 days ago
HTMLPaser	add some archive	6 months ago
JZSearch	update JZSearch in 2015/11/5	13 days ago
JZSearchclient	update JZSearch tools	14 days ago
KeyExtract	update some demos	2 months ago
NLPIR-ICTCLAS	update some demos	2 months ago
RedupRemover	update sample projects	3 months ago
SentimentAnalysis	update sentiment and SentimentAnalysis	2 months ago
SentimentNew	Update SentimentNew for both WIndows and Linux	2 months ago
Summary	update API, fix bug with NLPIR	2 months ago