

Engkoo

Parallel data mining from the web

Video:

<http://video.sina.com.cn/v/b/37417609-1286528122.html>

Rapidly Changing Language

- Approximately 1.5 billion people speak English as a primary, secondary or business language
- China: The largest “English speaking” country with 250 million English learners and USD 60 billion annual expenses
- Problem: Live language: new words, new meanings

Key Insight:

With billions of translated web pages and sharable repositories of language data growing every day, the Internet holds the sum of human language knowledge

www.engkoo.com

Microsoft Products:

Bing

词形变化 (无相关结果)

网络释义:

1. hidden rules

<http://61.152.116.30/bbs/read.php?tid=128369>

例句

类别: 全部

来源: [全部](#)

难度: [全部](#)

Human-Like TTS & Phonetic Search

4. Prev

对清创处理及二期创面处理进行

<http://www.ilib.cn/RSS-zndxxb-shkxb.xml>

5. Opening the Hidden Regulation of Officialism

破解官场潜规则

<http://www.iilib.cn/l-hebswdxzb.2006.02.html>

State-of-the-Art Machine Translation

(NIST OpenMT Winner)

Real-time Interactive Alignment

If we walk slow
如果我们比真

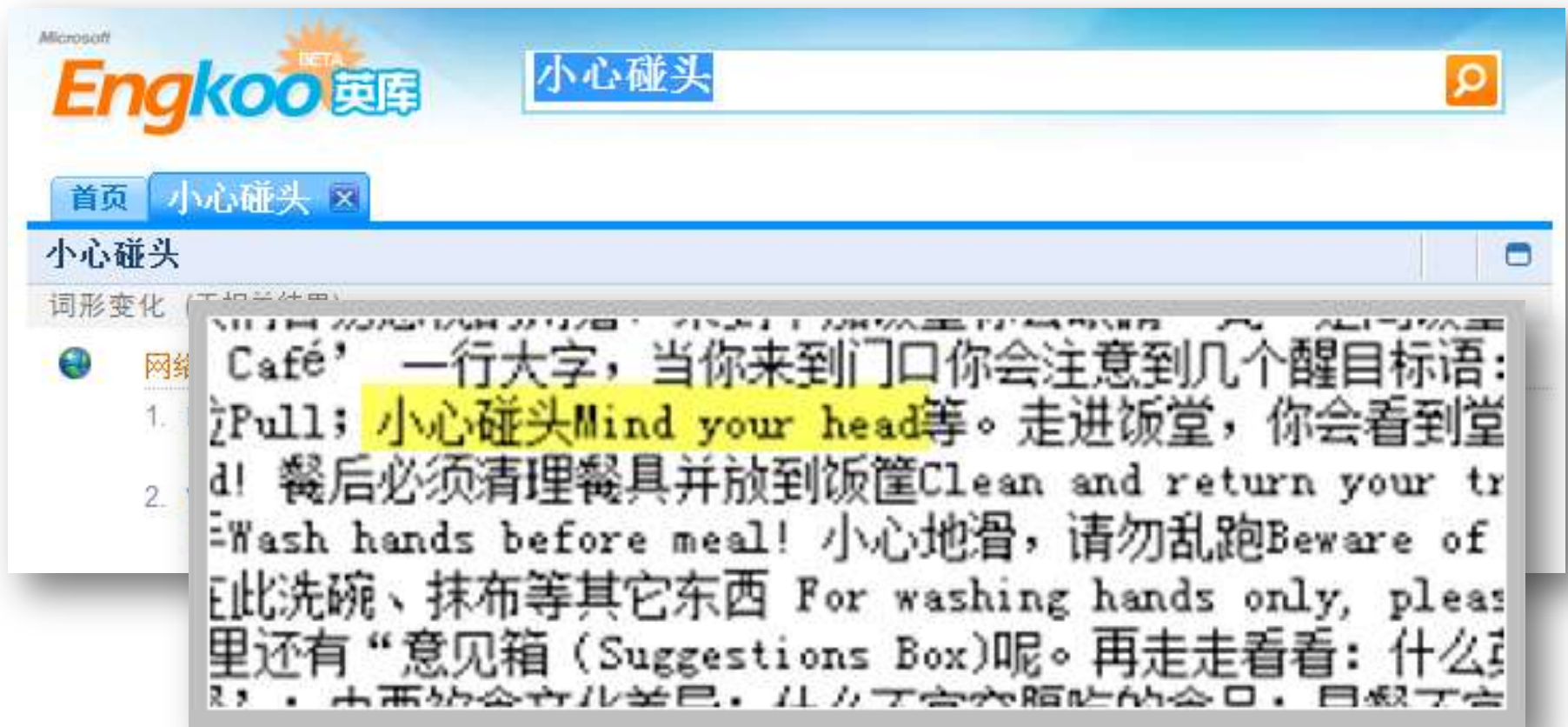
Office Jan

1 微软 2 微 3 喂 4 为 5 魏

MSN

[illegible]

Massive Dictionary Mined from the Web



Fresh and Diverse Examples

 BETA 英库

与时俱进

首页 例句:与时俱进

例句 类别: 全部 来源: 全部 难度: 全部

1. Go **ahead** with **times** and open up the new phase of community health ser
与时俱进开创社区卫生服务新局面
<http://www.ilib.cn/l-zgfybj.2003.11.html>
2. Develop **with** the **advance** of the **times** 
与时俱进
<http://218.22.70.71/printtext.asp?id=62917>
3. **Keeping pace** with the **times** and the Party mind line 
与时俱进与党的思想路线
<http://www.ilib.cn/l-hbdxxb-zxsh.2003.01.html>
4. Marx **and** Engels are the models of **keeping pace** with the **times** 
马克思恩格斯是与时俱进的典范

Advanced Search with Sentence Analysis

Microsoft
Engkoo 英库 BETA

Search bar: She is a **adj.** lady

Navigation: 首页 | 小心碰头 | She is a adj. l... x

Filters: 例句 | 类别: 全部 | 来源: 全部 | 难度: 全部 | ☐ 逐词释义

2. She is a **charming** young lady.
她是一个迷人的年轻女士。
<http://news.sogou.com/news?query=she%27s+a+lady>
3. She is a **perfect** lady.
她是一位十足的淑女。
<http://dict.bitunion.org/xdict.php?word=lady>
4. She is a **devout** old lady.
她是个虔诚的老太太。
<http://gz1.tougao.com/UserWork/alluser/901382/index/Article.asp?ArticleID=121507>

Microsoft

Engkoo 英库

微软Bing在美国的份额继续上升



首页

潜规则

微软Bing在美国... x

微软Bing在美国的份额继续上升

[报告问题或瑕疵](#)



计算机翻译:

Microsoft Bing shares continue to rise in the United States

微软Bing在美国的份额继续上升

Sentences Classification

例句 类别: 书面语 来源: 全部 难度: *****

1. The **mouse** sped down the oak-tree. 耗子刷地一下便下了橡树。
2. The **mouse** squeaked in the corner. 墙角处有老鼠的吱吱声。
3. They can grow in all **mouse** strain. 它们可以在所有的小鼠株内生长。

例句 类别: 技术 来源: 全部 难度: *****

1. That represents the **mouse** cursor. 它代表鼠标光标。
<http://msdn2.microsoft.com/en-us/library/system.windows.forms.c>
2. The user presses the **mouse** button. 用户按鼠标按钮。
<http://msdn2.microsoft.com/en-us/library/ms171542.aspx>
3. The user releases the **mouse** button. 用户释放鼠标按钮。



首页

mouse x

例句

类别:

全部

来源:

全部

难度:

全部

☒ 逐词释义

1. The **mouse** brushed down the oak-tree.
耗子刷地下了橡树。
2. The **mouse** squeaked in the corner.
墙角处有老鼠的吱吱声。
3. The **mouse** whisked into its hole.
这只老鼠急速地跑进洞去。
4. They can grow in all **mouse** strain.
它们可以在所有的小鼠株内生长。
5. The **mouse** looked at her rather inquisitively.
那只耗子用疑问的眼光看看她。
6. The girl shrieked when she saw the **mouse**.
那个姑娘一见耗子就吓得尖叫起来。
7. The **mouse** ran nimbly up the horse's leg.
那耗子灵活地顺着马腿爬上去。
8. "Ahem!" said the **Mouse** with an important air.



首页

mouse

例句

类别:

技术

来源:

全部

难度:

全部

☒ 逐词释义

1. That represents the **mouse** cursor.

它代表鼠标光标。

<http://msdn2.microsoft.com/en-us/library/system.windows.forms.cursor.current.aspx>

2. The user presses the **mouse** button.

用户按鼠标按钮。

<http://msdn2.microsoft.com/en-us/library/ms171542.aspx>

3. The user releases the **mouse** button.

用户释放鼠标按钮。

<http://msdn2.microsoft.com/en-us/library/ms171542.aspx>

4. Release the **mouse** button to drop the text.

释放鼠标按钮以放下文本。

[http://msdn2.microsoft.com/en-us/library/kddy153a\(VS.80\).aspx](http://msdn2.microsoft.com/en-us/library/kddy153a(VS.80).aspx)

5. A Microsoft **mouse** or compatible pointing device is required.

需要Microsoft鼠标或兼容的指点设备。

<http://msdn2.microsoft.com/en-us/library/ms143506.aspx>

6. The controllable **mouse** carries on the image programming demonstration,

可控鼠标进行图象编程示例

<http://www.88pl.com/soft/codelist.asp?page=324>

Learn Contextual Usage with Word Alignment

每日一句

If we walk slower **than** the others we shall **lag behind** them. 

如果我们**比**其他人走得慢,我们就会**落在**他们**后面**。



Learn Contextual Usage with Word Alignment

每日一句

If we walk slower than the others we shall lag behind them. 

如果我们比其他人走得慢,我们就会落在他们后面。



Learn Contextual Usage with Word Alignment

每日一句

If we **walk** slower than the others we shall **lag behind** them. 

如果我们比其他人**走**得慢,我们就会**落在**他们**后面**。



Hints of Easy-Confused Words



您要找的是不是:



音近词

- 🔊 physics
- 🔊 physical
- 🔊 physique
- 🔊 phoenix
- 🔊 Felix



形近词

- fizzes
- felix
- faikes
- finis
- fitios

Knowledge Mining Process

Parallel Sentence:

He could hardly afford to waste that golden time.
他无法浪费那样的好时光。

tokenizing: he could hardly afford to waste that golden time.
他无法浪费那样的好时光。

skeleton parsing: (Tsub~he~afford) (ModAdv~hardly) (Tobj~time~waste) (AdjAttrib~golden) (ModAdv~无法) (Tobj~他~浪费)
afford to(无法)

1. word's idiomatic usage

- Verb~Noun (decline~offer)
- Verb~Adv (greatly~improve)
- Adj~Noun (arduous~task)
- Adv~Adj (extremely~bad)

2. paraphrasing

- turn_on~light, switch_on~light
- laborious~task, hard~task
- deeply~moved, deeply~touched

1. single word

"he", "could", "hardly", "afford" etc.

"他", "无法", "浪费" etc.

2. single word with its POS

"he_Pron", "could_Verb", "hardly_Adv" etc.

"他_Pron", "无法_Adv", "浪费_Verb" etc.

3. collocation

"Tsub~he~afford", "Tobj~time~waste" etc.

"Tsub~他~浪费", "ModAdv~无法~浪费" etc.

Web Mining

Mined Data

Linguistic Parsing

Parsed Data

Linguistic Knowledge

Knowledge Mining

Multi-level Indexing

WSJ Asian Innovation Award: Reader's Choice Award

Microsoft Research
微软亚洲研究院

2010 Asian Innovation Awards

THE WALL STREET JOURNAL

The Wall Street Journal Asia recently announced the finalists for the 2010 Asian Innovation Awards. Out of close to 300 entries, submitted by various companies from all across Asia developing such things as language software and environmentally friendly technologies, Engkoo was selected as one of the 12 finalists, with a wider recognition for "helping Chinese internet users learn English as they surf."

Engkoo 英酷 Engkoo is a new kind of language assistance technology that enables Chinese people to ultimately master English as a native speaker right off online human translator mined from the web, machine translation, and a language learning experience, into one easy to use search and explore interface. By continuously discovering and incorporating high quality translation knowledge found on the web, this technology can be used to close the ever expanding translation gap between English and Chinese. In future work, the project will be extended to other language pairs.

THE WALL STREET JOURNAL
ASIAN INNOVATION AWARDS

Companies compete for 2010 prizes for technological advances in Asia

Products range from language software to more-efficient use of energy turning waste into building materials

Winning, whether you're one of the winners and you'll be competing for the prize, or you're one of the losers, it's a tough choice to make. The winners of the 2010 Asian Innovation Awards are now being announced. The winners of the 2010 Asian Innovation Awards are now being announced. The winners of the 2010 Asian Innovation Awards are now being announced.

Asian Innovation Awards 2010 Finalists

| Category | Winner | Runner-up |
|-------------------------|-----------------|-------------------|
| Artificial Intelligence | Microsoft | Engkoo |
| Biotechnology | Genentech | Novartis |
| Chemical | Roche | Novartis |
| Computer Hardware | Intel | AMD |
| Computer Software | Microsoft | Oracle |
| Consumer Electronics | Apple | Samsung |
| Energy | GE | Siemens |
| Environmental | Google | Facebook |
| Finance | JP Morgan Chase | Bank of America |
| Food & Beverage | Unilever | Nestle |
| Healthcare | Novartis | Roche |
| Internet | Google | Facebook |
| Manufacturing | Toyota | Honda |
| Materials | 3M | DuPont |
| Medical Devices | Medtronic | Johnson & Johnson |
| Mobile | Apple | Samsung |
| Robotics | ABB | KUKA |
| Space | SpaceX | Boeing |
| Transportation | Boeing | Airbus |
| Utilities | GE | Siemens |



Bilingual Data (BD) in Engkoo

Microsoft
Engkoo 英库 BETA

奥巴马

首页 奥巴马

奥巴马

词形变化 (无相关结果)

网络释义:

1. barack obama
<http://bbs.haozhai.com/viewthread.php?tid=94027>

例句

类别: 全部 来源: 全部 难度: 全部 逐词释义

1. The man handed me a pamphlet. "Mr. Obama, I want you to know that I agree with a lot of what you have to say."
那个男人递给了我一本小册子并说道,"奥巴马先生,我想告诉你,我并不同意你要说的".
http://groups.google.com/group/mimiqiao/browse_thread/thread/19080e4523a27d15/4d7a471...
2. Obama, Romney to Formally Enter US Presidential Race
奥巴马和罗姆尼正式宣布参加下届总统竞选
<http://board.verycd.com/t460479.html>

Mining BD from the Web

- From parallel pages
 - (Shi et al., ACL-06)
- From search results
 - (Jiang et al., IJCAI-07)
- Using parenthesis pattern
- From collective bilingual pages
 - (Jiang et al., ACL-09)

Parallel Page

<http://support.microsoft.com/kb/914962/en-us>

<http://support.microsoft.com/kb/914962/zh-cn>

This article lists problems that are fixed in Microsoft Windows Server 2003 Service Pack 2 (SP2). Service packs are cumulative. This means that the problems that are fixed in a service pack are also fixed in later service packs.

[↑ Back to the top](#)

This article contains a list of Microsoft Knowledge Base (KB) articles that contain updates that are contained in Microsoft Windows Server 2003 Service Pack 2 (SP2).

This article is primarily intended to help you troubleshoot problems with Windows Server 2003 SP2, the following answers:

<http://support.microsoft.com/ph/3198/>

本文列出了 Microsoft Windows Server 2003 Service Pack 2 (SP 2) 中修复的问题。Service Pack 具有累积性。这意味着在一个 Service Pack 中修复该问题还会在以后的 Service Pack 中修复。

[↑ 回到顶端](#)

本文描述的修复程序和 Microsoft Windows Server 2003 Service Pack 2 (SP 2) 中包含的更新的 Microsoft 知识库 (KB) 文章列表。

本文主要被为了帮助 IT 专业人员和支持和维护公司已经计算机系统上的公司 helpdesks。所有其他人对有疑问 Windows Server 2003 SP 2，下面的 Microsoft Web 站点可能是一个更好的站点，以查找答案：

<http://support.microsoft.com/ph/3198/>

Mining BD from Parallel Pages

Source



Tr
w

Web Dire

Web De

"Chinese v

+Search Bu

Web R

Page 1 of 12

bing

Discount

English V

www.gocam

Disco

Englis

Intern

www.g

Sho

Travel and your tips for China

... Airport Codes | Consolidated Airfare Quotes | Travel ... to find your way back as t
may be quite different from the Chinese version.

www.travel360degrees.com/article0022.htm Cached page

Web page

新渡輪
FIRST FERRY
兩大福企業及新創集團成員
Member of Chow Tai Fook Enterprises & NWS Holdings

Our Story

Our Story
Company News
Customer Service
Know More About NWS
First Ferry Club
Cyber Shop
Funtan
Hotspots
News Channels
Weather Forecast
Useful Links

Site Map

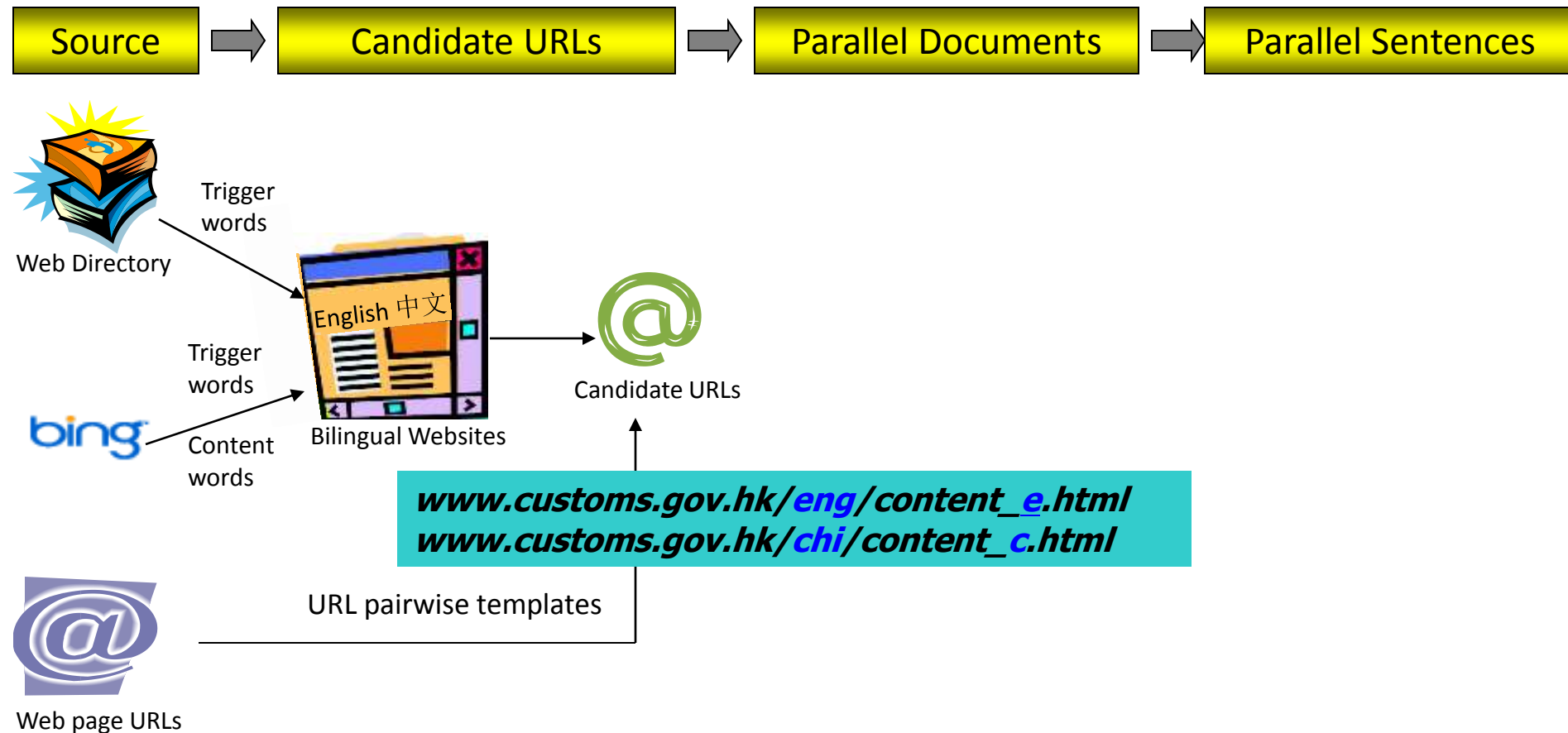
Our Story

New World First Ferry Services Limited is jointly owned by Chow Tai Fook Enterprises Limited and [NWS Holdings Limited](#) ("NWS Holdings"). First Ferry was established in November 1999, and commenced its service on 15 January 2000, and operates a total of eight outlying and inner harbour ferry routes at present.

Under the same umbrella, New World First Ferry Servicos Maritimos (Macau) Limitada [First Ferry (Macau)] has operated ferry service between Macau and Tsim Sha Tsui, Hong Kong since January 2000.

"Customer First and Foremost" is our corporate mission, First Ferry strives to provide a safe and comfortable journey to the commuting public. Quality people and a reliable vessel fleet are essences in our quest for quality ferry service.

Mining BD from Parallel Pages



Mining BD from Parallel Pages

COURT OF FINAL APPEAL

The Court of Final Appeal is the highest appellate court in the Hong Kong Special Administrative Region. It has jurisdiction in respect of matters conferred on it by the **Hong Kong Court of Final Appeal Ordinance**, Cap. 484 and by any other law.

It hears appeals on civil and criminal matters from the High Court (the Court of Appeal and the Court of First Instance).

Parallel Sentences

Trigger

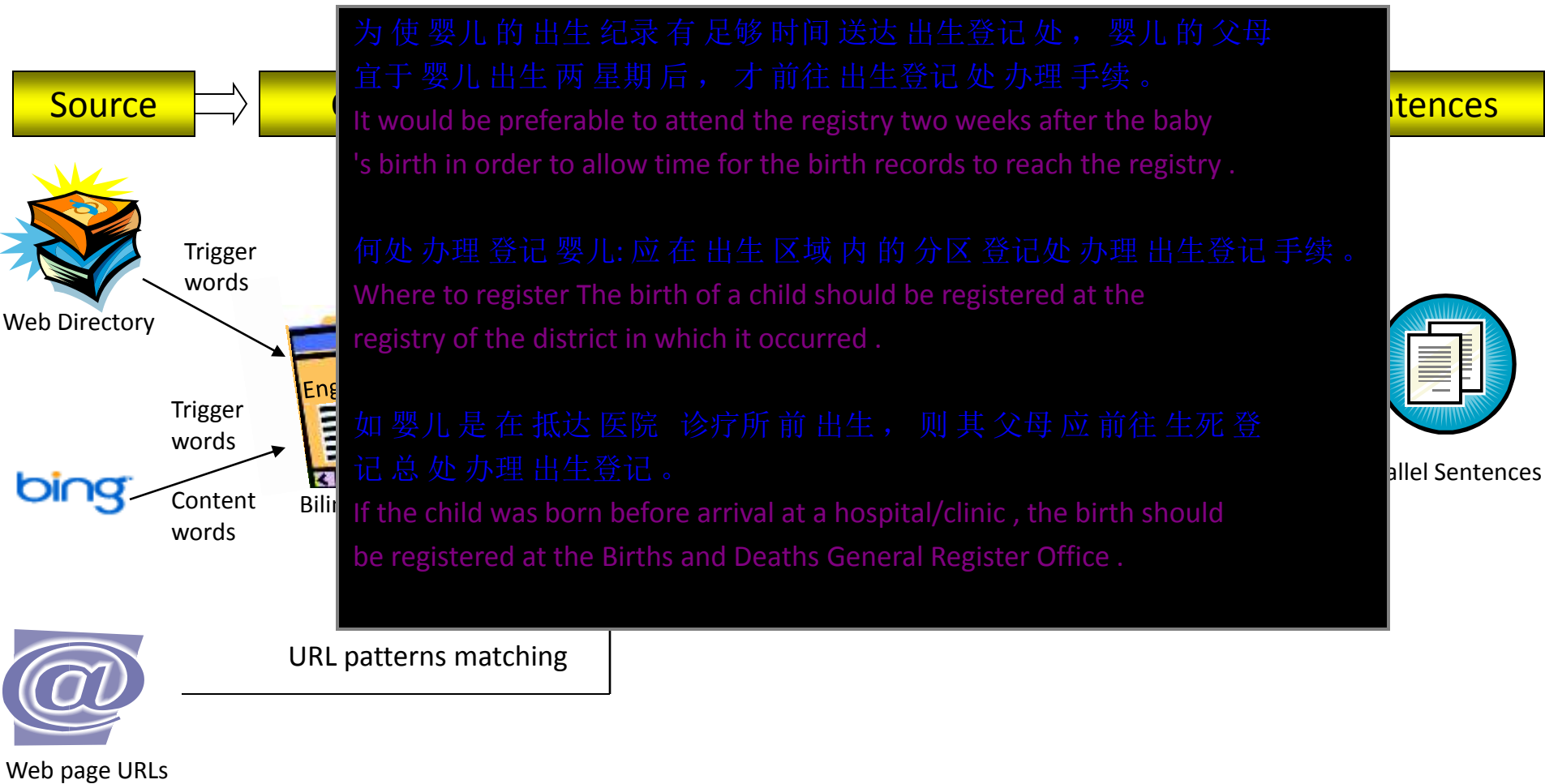


终审法院

终审法院是香港特别行政区最高的上诉法院，根据香港法例 第484章《香港终审法院条例》赋予的权力，处理针对高等法院（上诉法庭及原讼法庭）的判决而作出的上诉及有关事项。

Parallel hyperlinks

Mining BD from Parallel Pages



Mining BD from Search Results

bing Beta

ALL RESULTS

Microsoft Research Asia

Search for Chinese pages

Show all Only English Your settings

ALL RESULTS

1-10 of 26,800 results - [Advanced](#)

[Microsoft Research Asia](#) [Translate this page](#)
[igroup.msra.cn](#) - [Mark as spam](#)

[微软亚洲研究院](#) [Translate this page](#)
Microsoft Research Asia English Website | 网站地图 最新发布 创新成果 热门下载
[www.msra.cn](#) - [Cached page](#) - [Mark as spam](#)

[微软亚洲研究院](#) [Translate this page](#)
[微软亚洲研究院](#) **Microsoft Research Asia**
[www.msra.cn/rmc](#) - [Cached page](#) - [Mark as spam](#)

[Microsoft Research Asia Theory Workshop](#) [Translate this page](#)
Organizers Andrej Bogdanov (Tsinghua University, andrejb@tsinghua.edu.cn) Dr. Wei Chen (MSR Asia, weic@microsoft.com) Cynthia Dwork (MSR Silicon Valley, dwork@microsoft.com) Shanghua Teng (Boston University / MSR Asia, steng@cs.bu.edu) The goal of the workshop is to enhance the communication between the nascent ...
[research.microsoft.com/en-us/um/beijing/events/theory...](#) - [Cached page](#) - [Mark as spam](#)

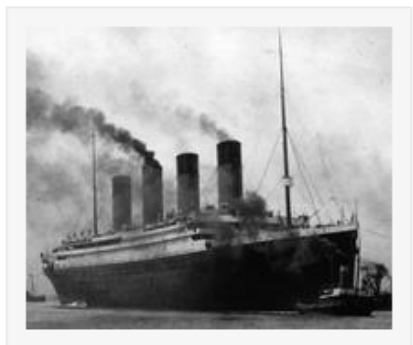
[Microsoft Research Asia Family Day - Windows Live](#) [Translate this page](#)
June 04 **Microsoft Research Asia Family Day** 上个月的wine down,公司把它推迟到今天, 作为一次儿童节聚会, 邀请公司的成员带上自己的家属, 特别是那些有孩子的员工, 让孩子们和爸爸妈妈一起开心一下 :) 记个流水账//
[cloudsun.spaces.live.com/blog/cns!927F11B12223353A!260.entry](#) - [Cached page](#) - [Mark as spam](#)

Mining BD using Parenthesis Pattern

泰坦尼克号邮船简介

[编辑本段]

英国皇家邮船泰坦尼克号(RMS Titanic)是奥林匹克级邮轮的第二艘邮船，20世纪初，由英国白星航运公司(White Star Line)制造的一艘巨大豪华客轮。由位于爱尔兰贝尔法斯特(Belfast)的哈兰德与沃尔夫(Harland and Wolff)造船厂兴建。泰坦尼克号是当时世界上最大的豪华客轮，被称为是“永不沉没的船”或是“梦幻之船”。泰坦尼克号共耗资7500万英镑，吨位46328吨，长882.9英尺，宽92.5英尺，从龙骨到四个大烟囱的顶端有175英尺，高度相当于11层楼，是当时一流的超级豪华巨轮。计划与姐妹船奥林匹克号(RMS Olympic)和不列颠尼克号(RMS Britannic)一道为英国白星航运公司的乘客们提供快速且舒适的跨大西洋旅行。

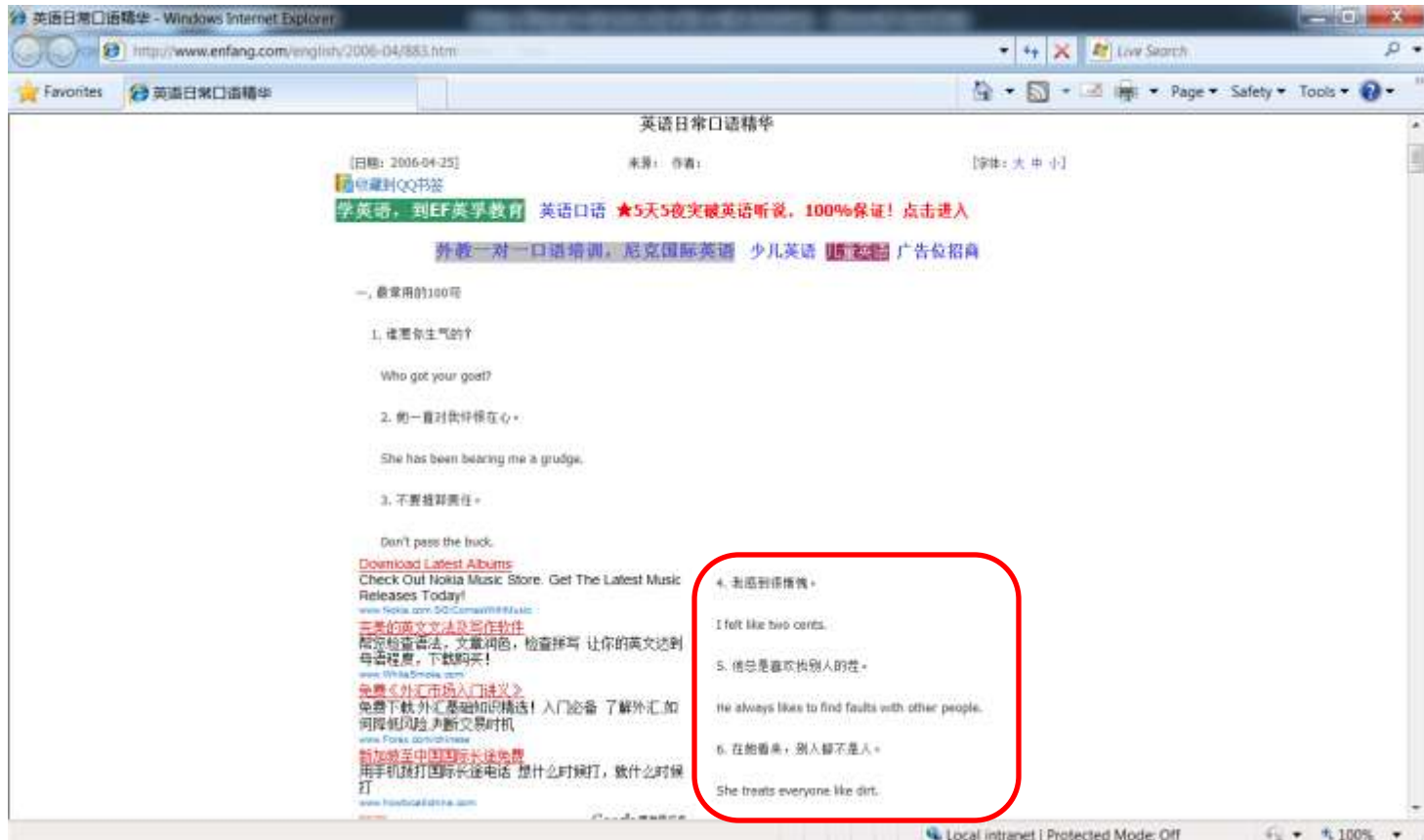


难而改编。

1912年4月10日，泰坦尼克号从英国南安普敦(Southampton)出发，途经法国瑟堡-奥克特维尔(Cherbourg-Octeville)以及爱尔兰昆士敦(Queenstown)，计划中的目的地为美国的(New York)，开始了这艘“梦幻客轮”的处女航。4月14日晚11点40分，泰坦尼克号在北大西洋撞上冰山(大约在41°43'55.66"N 49°56'45.02"W附近)，两小时四十分钟后，4月15日凌晨2点20分沉没，由于缺少足够的救生艇，1500人葬生海底，造成了当时在和平时期最严重的一次航海事故，也是迄今为止最为人所知的一次海难。电影《泰坦尼克号》就是根据这一真实海

Mining BD from Collective Bilingual Pages

- Collective Bilingual Page (CBP): pages in which bilingual data **appear collectively** and are **formatted consistently**



Mining BD from Collective Bilingual Pages

- Collective Bilingual Pages
collectively

4. 我感到很惭愧。

I felt like two cents.

5. 他总是喜欢找别人的茬。

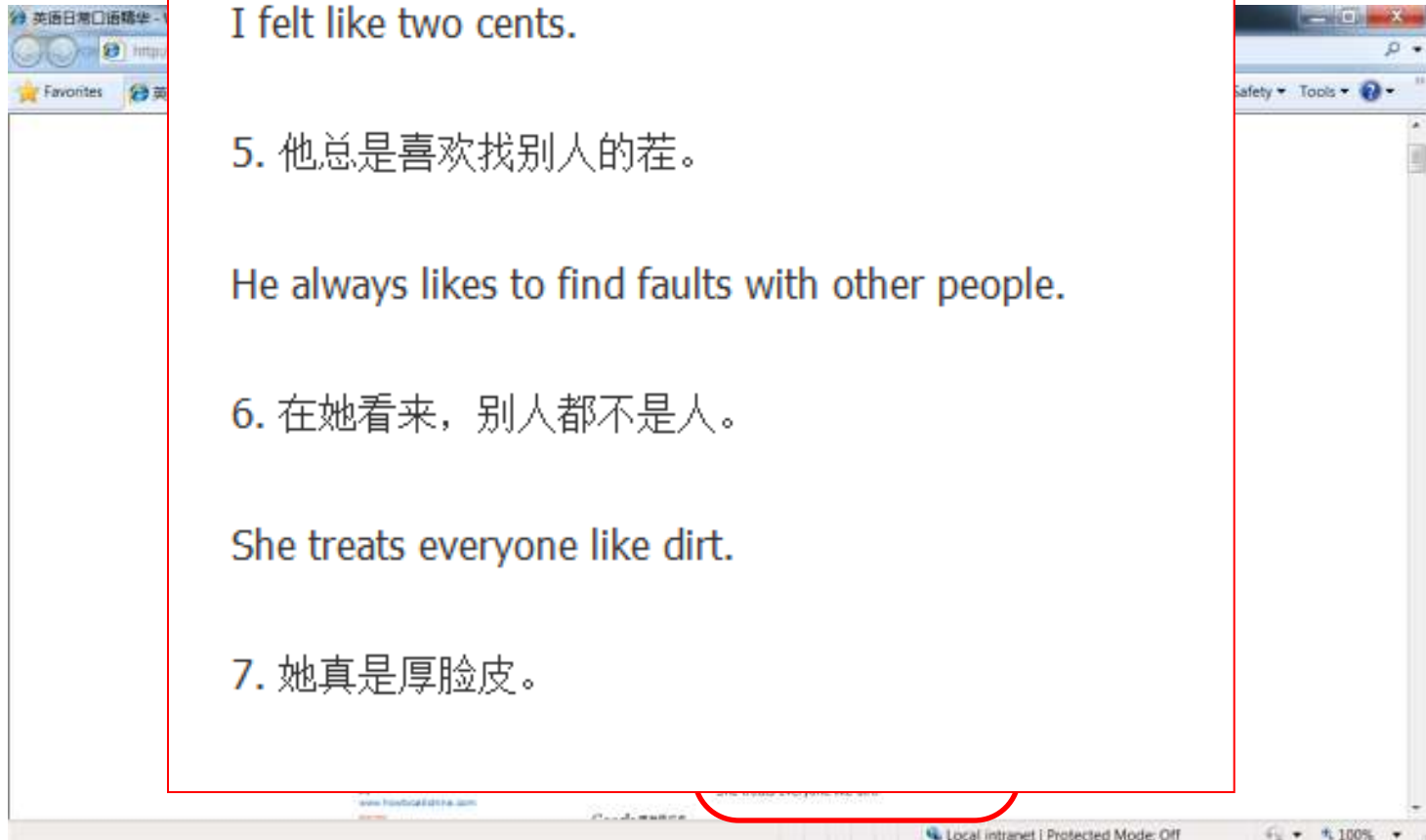
He always likes to find faults with other people.

6. 在她看来，别人都不是人。

She treats everyone like dirt.

7. 她真是厚脸皮。

appear



Mining BD from Collective Bilingual Pages

- Collective Bilingual Page (CBP): pages in which bilingual data **appear collectively** and are **formatted consistently**

工作犬组 (Working Group) 的全部品种的正确名称和分类 - Windows Internet Explorer

http://www.chinapet.net/db/viewtopic.php?t=256115&bbiddata=old

Favorite 工作犬组 (Working Group) 的全部品种

ray7501
宠物大帝

注册时间: 2004-02-14
帖子: 1503
来自: k9china.com
927.6 金币

发表于: 2005-04-24 09:56 发表主题: Re: 工作犬组 (Working Group) 的全部品种的正确名称和分类!

干美 写到:

1. Alaskan Malamute 阿拉斯加雪橇犬
2. Beauceron 法国灰犬
3. Bernese Mountain Dog 伯恩山犬
4. Bouvier des Flandres 比利牛斯牛犬
5. Boxer 拳师
6. Bullmastiff 斗牛獒
7. Cane Corso 卡斯罗
8. Doberman 杜宾
9. Dogue de Bordeaux 波多尔
10. Eskimo Dog 爱斯基摩犬
11. German Pinscher 德国梗毛草犬
12. Giant Schnauzer 大型雪纳瑞
13. Great Dane 大丹
14. Hovawart 霍夫瓦尔特犬
15. Leonberger 德国莱昂伯格犬
16. Mastiff 马士提夫
17. Neapolitan Mastiff 拿波里獒犬
18. Newfoundland 纽芬兰
19. Portuguese Water Dog 葡萄牙水犬
20. Rottweiler 罗威那
21. Russian Black Terrier 黑俄罗斯梗
22. Siberian Husky 西伯利亚雪橇犬
23. South African Boerboels
24. St. Bernard 圣伯纳
25. Tibetan Mastiff 西藏獒犬

South African Boerboels

sa_boerboel.jpg


| | |
|-----|----------|
| 描述: | |
| 大小: | 24.71 KB |

Done

Local intranet | Protected Mode: Off

Mining BD from Collective Bilingual Pages

- Collective Bilingual Page (CBP): pages in which bilingual data **appear collectively** and are **formatted consistently**



The screenshot shows a web browser window with a list of 11 dog breeds. Each entry consists of a number, the breed name in English, and its Chinese name. The list is as follows:

1. Alaskan Malamute 阿拉斯加雪橇犬
2. Beauceron 法国狼犬
3. Bernese Mountain Dog 伯恩山地犬
4. Bouvier des Flandres 比利时牧牛犬
5. Boxer 拳师
6. Bullmastiff 斗牛獒
7. Cane Corso 卡斯罗
8. Dobermann 杜宾
9. Dogue de Bordeaux 波多尔
10. Eskimo Dog 爱斯基摩犬
11. German Pinscher 德国硬毛宾莎犬

On the left side of the browser window, there is a sidebar with a user profile for 'ray7501' (宠物大帝) and a small image of a dog. Below the profile, it shows registration information: '注册时间: 2004-02-14', '帖子: 1503', '来自: k9china.com', and '927.6 金币'. At the bottom of the browser window, there is a status bar showing 'Done' and 'Local intranet | Protected Mode: Off'.

Mining Bilingual Data from CBPs

- Identification of seed translation pairs

1. Alaskan Malamute 啊拉斯加雪橇犬
2. Beauceron 法国狼犬
3. **Bernese Mountain Dog 伯恩山地犬**
4. Bouvier des Flandres 比利时牧牛犬
5. **Boxer 拳师**
6. Bullmastiff 斗牛犬
7. Cane Corso 坎科索犬
8. Dobermann 杜宾犬
9. Dogue de Bordeaux 波多斯犬
10. **Eskimo Dog 爱斯基摩犬**
11. German Pinscher 德国 pinscher 犬

Bernese Mountain Dog
伯恩 山地 犬

Mining Bilingual Data from CBPs

- Pattern: <number> <period> <English-term> <Chinese-term>

| | | |
|-----|----------------------|---------|
| 1. | Alaskan Malamute | 啊拉斯加雪橇犬 |
| 2. | Beauceron | 法国狼犬 |
| 3. | Bernese Mountain Dog | 伯恩山地犬 |
| 4. | Bouvier des Flandres | 比利时牧牛犬 |
| 5. | Boxer | 拳师 |
| 6. | Bullmastiff | 斗牛獒 |
| 7. | Cane Corso | 卡斯罗 |
| 8. | Dobermann | 杜宾 |
| 9. | Dogue de Bordeaux | 波多尔 |
| 10. | Eskimo Dog | 爱斯基摩犬 |
| 11. | German Pinscher | 德国硬毛宾莎犬 |

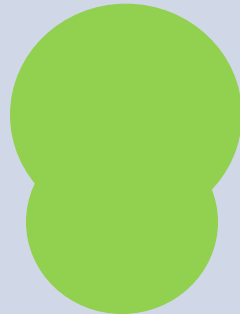
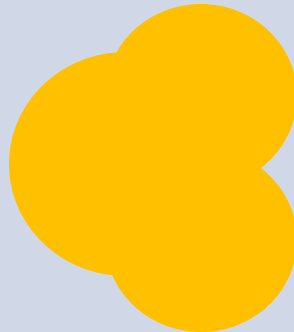
Mining Bilingual Data from CBPs

- Pattern: <number> <period> <English-term> <Chinese-term>

| | | |
|-----|----------------------|---------|
| 1. | Alaskan Malamute | 啊拉斯加雪橇犬 |
| 2. | Beauceron | 法国狼犬 |
| 3. | Bernese Mountain Dog | 伯恩山地犬 |
| 4. | Bouvier des Flandres | 比利时牧牛犬 |
| 5. | Boxer | 拳师 |
| 6. | Bullmastiff | 斗牛獒 |
| 7. | Cane Corso | 卡斯罗 |
| 8. | Dobermann | 杜宾 |
| 9. | Dogue de Bordeaux | 波多尔 |
| 10. | Eskimo Dog | 爱斯基摩犬 |
| 11. | German Pinscher | 德国硬毛宾莎犬 |

Results of BD Mining

- From parallel pages
 - (Shi et al., ACL-06)
 - About 6 million bilingual sentences
- From search results
 - (Jiang et al., IJCAI-07)
 - About 6 million bilingual terms
- Using parenthesis pattern
 - About 6 million bilingual terms
- From collective bilingual pages
 - (Jiang et al., ACL-09)
 - About 10.5 million bilingual sentences and 10.1 million bilingual terms

| Bilingual sentences | Bilingual terms |
|---|---|
|  |  |
| 14 million | 17 million |

Summary

- Mine massive translation knowledge from the web to build a huge lexicon
- Build a search engine to provide dictionary explanation and example sentences
- Learn a SMT engine with the mined example sentences for translation references
- Apply NLP based search to enable advanced search
- Learn pronunciation with TTS
- Language-independent technologies extensible to other language pairs