
University of Glasgow
工程与科研导论
INSPIRING PEOPLE

工程

科研

Web工程

前端

HTML, CSS, Javascript

后端

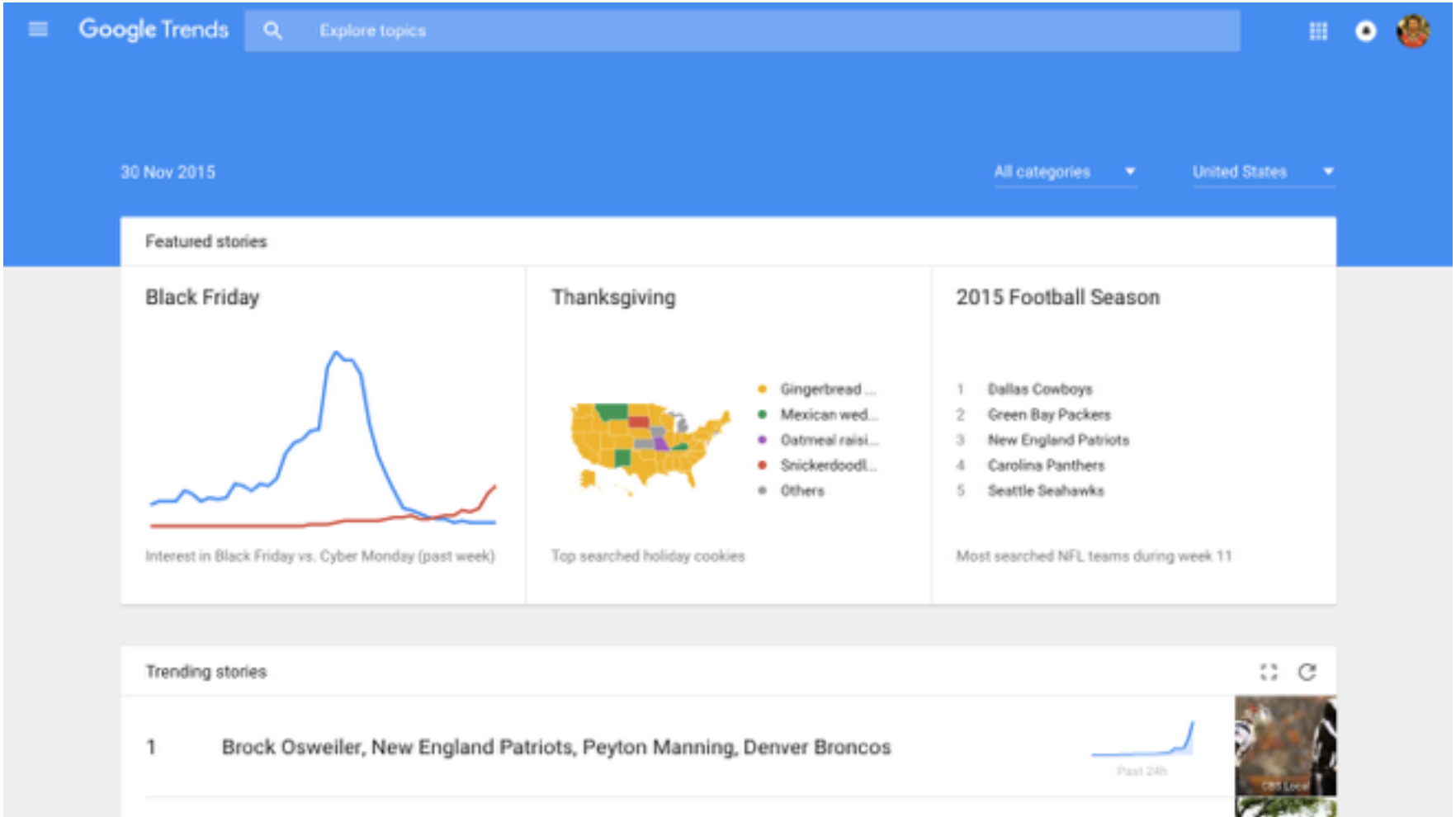
PHP, Node.js, Golang等

系统工程

C/C++等

Linux

Google Trends(Google趋势)

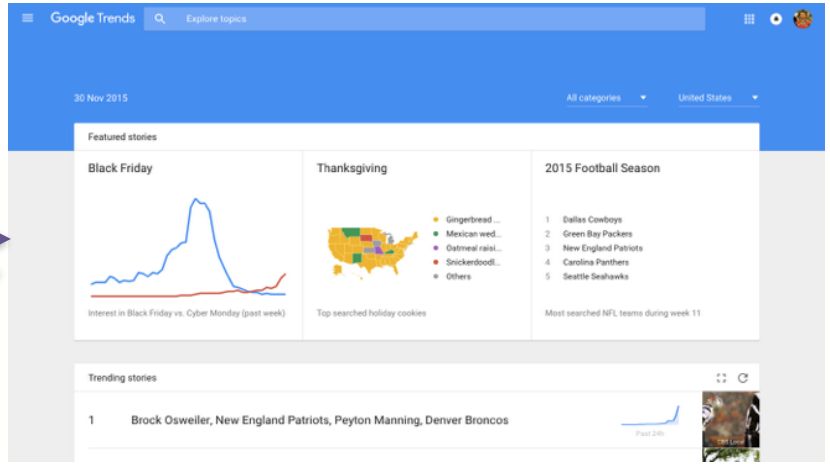


Web工程

PHP, Node.js, Golang

HTML
CSS
Javascript

数据

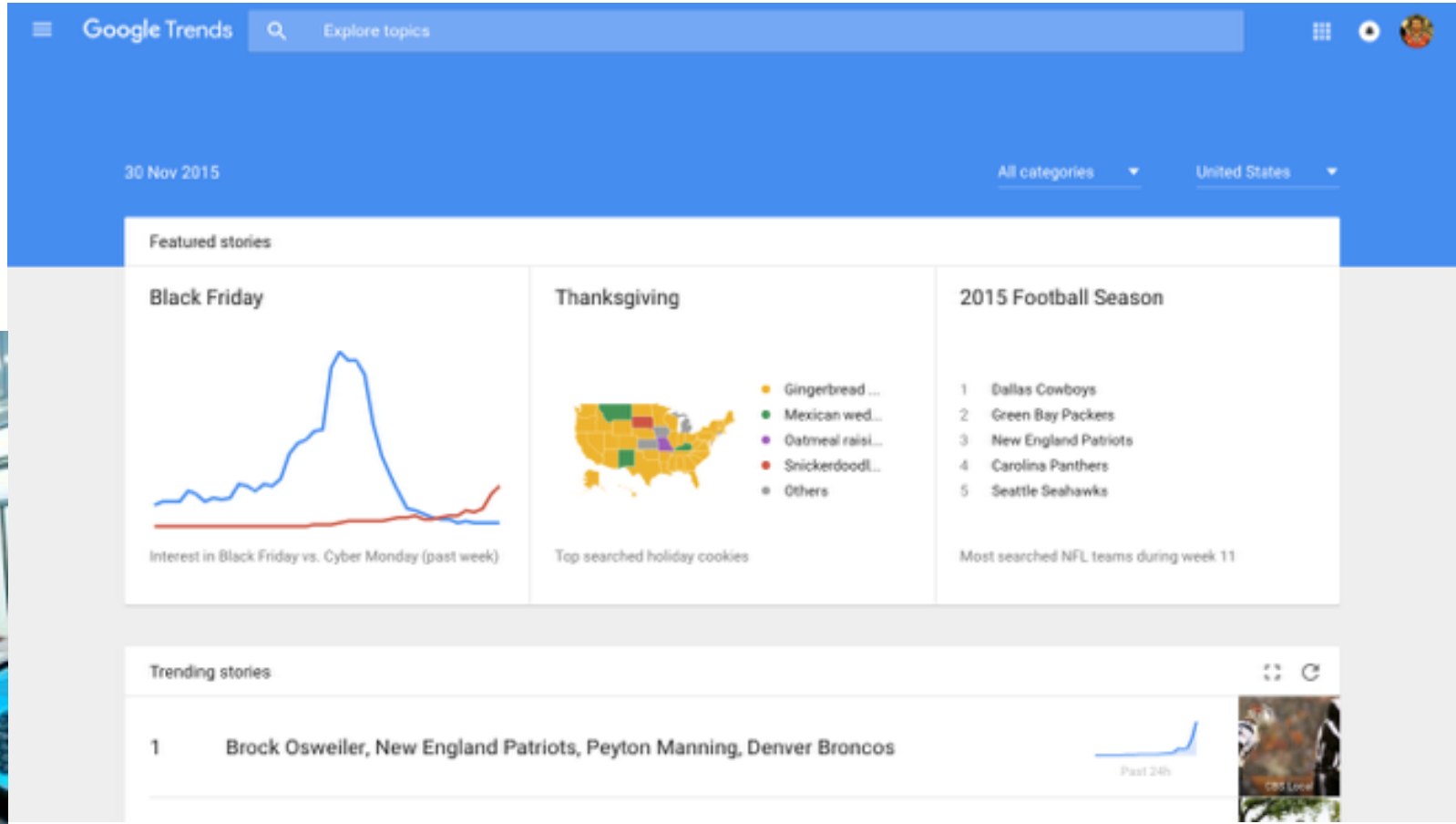


後端

前端

前端

HTML
CSS
Javascript



科研

- 背景研究（Background Survey）
- 研究方法、模型（Methodology）
- 实验分析（Experiment/Evaluation）
  - Baseline
  - State-of-the-art
  - Your Model

简介

- 背景研究
- 数据准备
- 数学模型

背景研究

只有知道当前研究的现状，
才会做出有意义的科研。

背景研究


中国计算机学会推荐国际学术刊物

（数据库、数据挖掘与内容检索）

一、A类

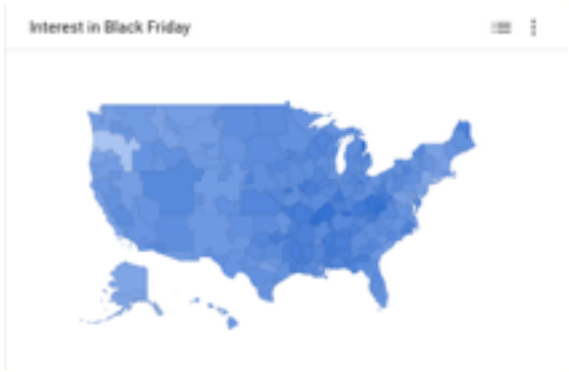
序号	刊物简称	刊物全称	出版社	网址
1	TODS	ACM Transactions on Database Systems	ACM	http://www.acm.org/tods/
2	TOIS	ACM Transactions on Information and Systems	ACM	http://www.acm.org/pubs/tois/
3	TKDE	IEEE Transactions on Knowledge and Data Engineering	IEEE Computer Society	http://www.computer.org/tkde/
4	VLDBJ	VLDB Journal	Springer-Verlag	http://www.vldb.org/dblp/db/journals/vldb/index.html

二、B类

序号	刊物简称	刊物全称	出版社	网址
1	TKDD	ACM Transactions on Knowledge Discovery from Data	ACM	http://www.acm.org/pubs/tkdd/
2	AEI	Advanced Engineering Informatics	Elsevier	http://www.elsevier.com/wps/find/journaldescription.cws_home/622240/
3	DKE	Data and Knowledge Engineering	Elsevier	http://www.sciencedirect.com/science/journal/0169023X
4	DMKD	Data Mining and Knowledge Discovery	Springer	http://www.springerlink.com/content/100234/
5	EJIS	European Journal of Information Systems	The OR Society	http://www.palgrave-journals.com/ejis/
6		GeoInformatica	Springer	http://www.springerlink.com/content/1573-7624/
7	IPM	Information Processing and Management	Elsevier	http://www.elsevier.com/locate/infoproman
8		Information Sciences	Elsevier	http://www.elsevier.com/locate/issn/00200255
9	IS	Information Systems	Elsevier	http://www.journals.elsevier.com/information-systems/

Peer-review论文审查

趋势变化图



如何获取到数据?

数据准备

1. API
2. 爬虫

Twitter API

https://dev.twitter.com/rest/public

REST APIs

The REST APIs provide programmatic access to read and write Twitter data. Author a new Tweet, read author profile and follower data, and more. The REST API identifies Twitter applications and users using OAuth; responses are available in JSON.

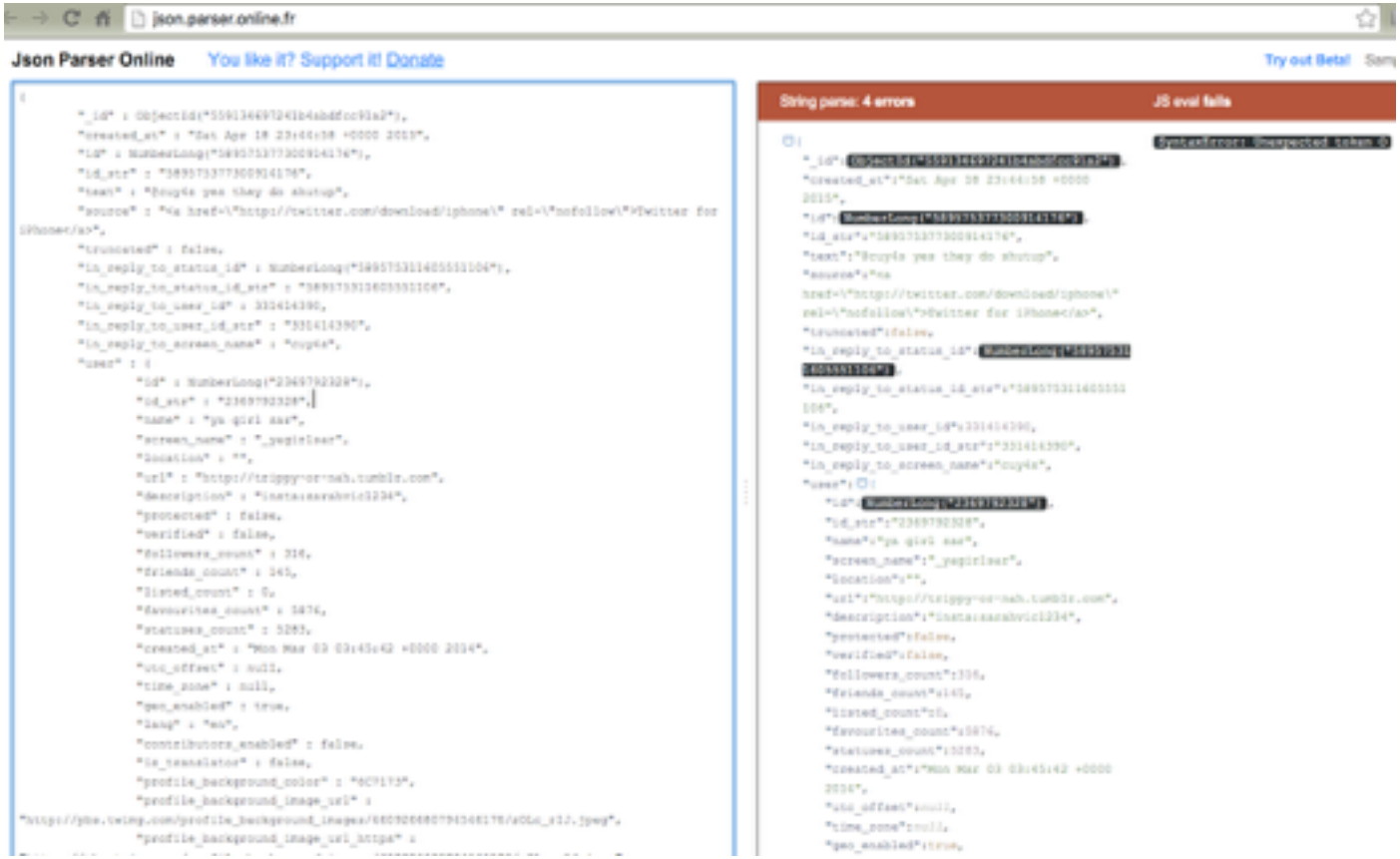
If your intention is to monitor or process Tweets in real-time, consider using the Streaming API instead.

Overview

Below are the documents that will help you get going with the REST APIs as quickly as possible

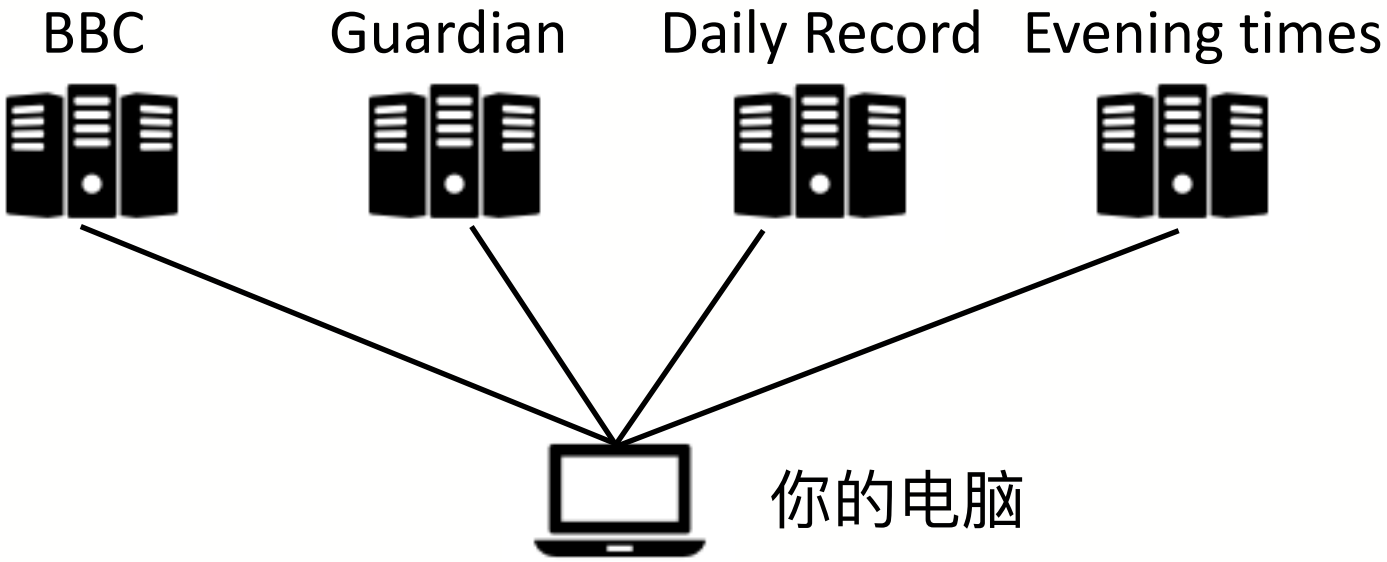
- API Rate Limiting
- API Rate Limits
- Working with Timelines
- Using the Twitter Search API
- Uploading Media
- Multiple Media Entities in Statuses
- Finding Tweets about Places

Twitter Data



Json Online(Json在线编辑器)

爬虫



爬虫

Get请求抓取页面(Python)

```
import urllib2
content = urllib2.urlopen('http://baidu.com').read()
print content
```


处理HTML页面

Python： Beautiful Soup

格式化数据

XML/JSON

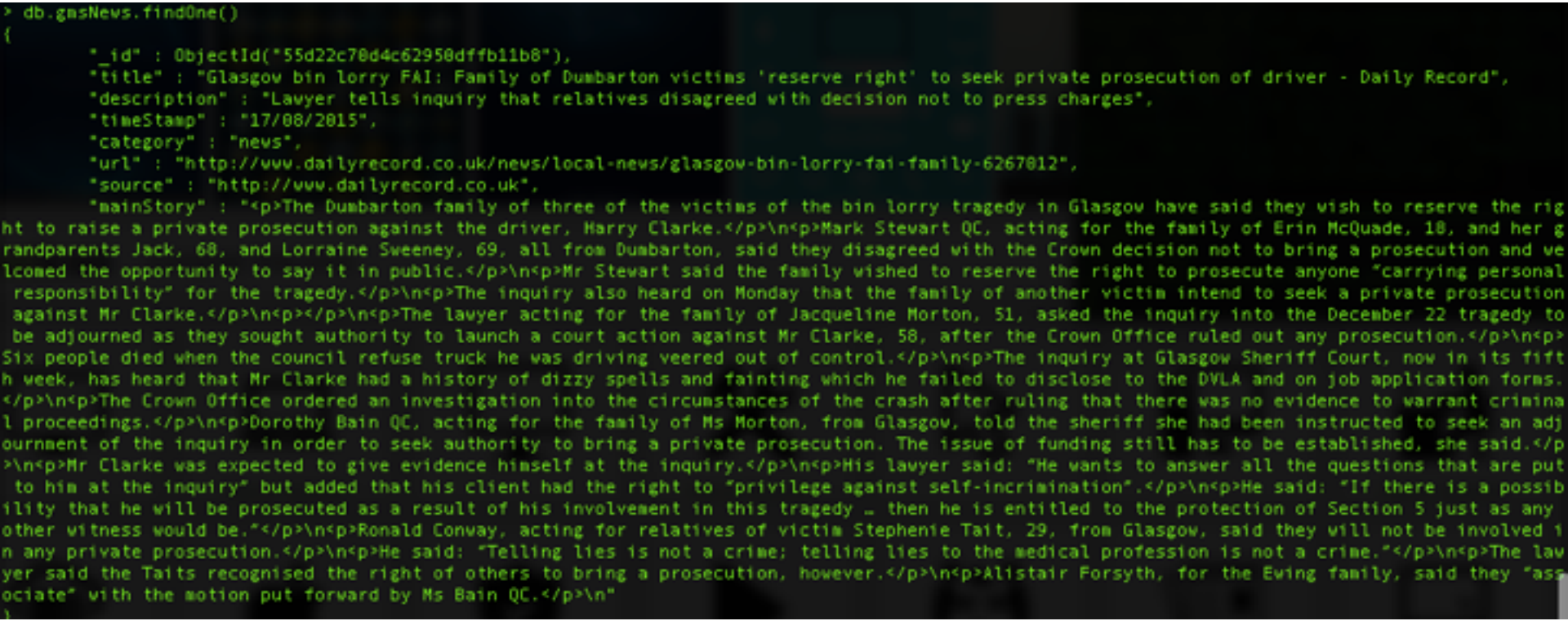
JSON优势:

MongoDB数据库的结合

和map类型的灵活转换

爬虫

MongoDB中的爬虫数据



```
> db.gmsNews.findOne()
{
        "_id" : ObjectId("55d22c70d4c62950dffb11b8"),
        "title" : "Glasgow bin lorry FAI: Family of Dumbarton victims 'reserve right' to seek private prosecution of driver - Daily Record",
        "description" : "Lawyer tells inquiry that relatives disagreed with decision not to press charges",
        "timeStamp" : "17/08/2015",
        "category" : "news",
        "url" : "http://www.dailyrecord.co.uk/news/local-news/glasgow-bin-lorry-fai-family-6267012",
        "source" : "http://www.dailyrecord.co.uk",
        "mainStory" : "<p>The Dumbarton family of three of the victims of the bin lorry tragedy in Glasgow have said they wish to reserve the rig
ht to raise a private prosecution against the driver, Harry Clarke.</p>\n<p>Mark Stewart QC, acting for the family of Erin McQuade, 10, and her g
randparents Jack, 68, and Lorraine Sweeney, 69, all from Dumbarton, said they disagreed with the Crown decision not to bring a prosecution and we
lcomed the opportunity to say it in public.</p>\n<p>Mr Stewart said the family wished to reserve the right to prosecute anyone "carrying personal
 responsibility" for the tragedy.</p>\n<p>The inquiry also heard on Monday that the family of another victim intend to seek a private prosecution
 against Mr Clarke.</p>\n<p></p>\n<p>The lawyer acting for the family of Jacqueline Morton, 51, asked the inquiry into the December 22 tragedy to
 be adjourned as they sought authority to launch a court action against Mr Clarke, 58, after the Crown Office ruled out any prosecution.</p>\n<p>
Six people died when the council refuse truck he was driving veered out of control.</p>\n<p>The inquiry at Glasgow Sheriff Court, now in its fift
h week, has heard that Mr Clarke had a history of dizzy spells and fainting which he failed to disclose to the DVLA and on job application forms.
</p>\n<p>The Crown Office ordered an investigation into the circumstances of the crash after ruling that there was no evidence to warrant crimina
l proceedings.</p>\n<p>Dorothy Bain QC, acting for the family of Ms Morton, from Glasgow, told the sheriff she had been instructed to seek an adj
ournment of the inquiry in order to seek authority to bring a private prosecution. The issue of funding still has to be established, she said.</p
>\n<p>Mr Clarke was expected to give evidence himself at the inquiry.</p>\n<p>His lawyer said: "He wants to answer all the questions that are put
 to him at the inquiry" but added that his client had the right to "privilege against self-incrimination".</p>\n<p>He said: "If there is a possib
ility that he will be prosecuted as a result of his involvement in this tragedy … then he is entitled to the protection of Section 5 just as any
other witness would be."</p>\n<p>Ronald Conway, acting for relatives of victim Stephenie Tait, 29, from Glasgow, said they will not be involved i
n any private prosecution.</p>\n<p>He said: "Telling lies is not a crime; telling lies to the medical profession is not a crime."</p>\n<p>The law
yer said the Taits recognised the right of others to bring a prosecution, however.</p>\n<p>Alistair Forsyth, for the Ewing family, said they "ass
ociate" with the motion put forward by Ms Bain QC.</p>\n"
}
```

数学模型

1. 机器学习

《Pattern Recognition and Machine Learning》
《Elements of Statistical Learning》(偏统计学)

A. 微积分
B. 概率论
C. 统计

2. Kaggle机器学习竞赛

www.kaggle.com

公式推导

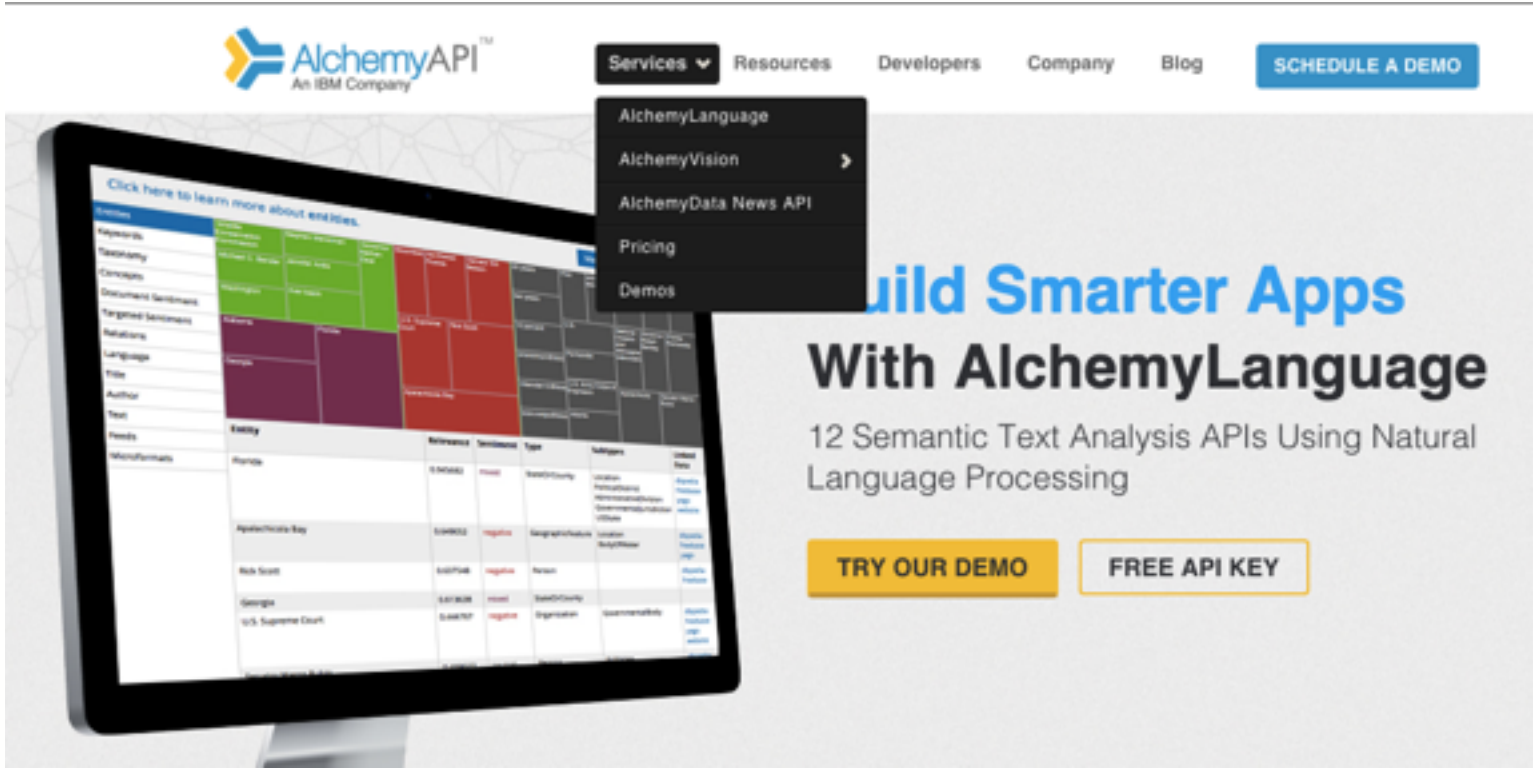
$$
\begin{aligned}
p(z, w^{(k)}|\mathcal{A}, \alpha^{(k)}, \beta) &= \int_x \int_\Phi \int_\mathcal{H} p(z|x,\Phi) p(w^{(k)}|z,\mathcal{H}) p(x|\mathcal{A}) p(\Phi|\alpha^{(k)}) p(\mathcal{H}|\beta) dx d\Phi d\mathcal{H} \\
&= \int_x \int_\Phi \int_\mathcal{H} \left[\prod_{i=1}^{N} p(z_{di}|\phi_{x_{di}})\right] \left[\prod_{d=1}^{D}\prod_{i=1}^{N_d} p(x_{di}|a_d)\right] p(\Phi|\alpha^{(k)}) p(\mathcal{H}|\beta) dx d\Phi d\mathcal{H} \\
&= \int_x \int_\Phi \int_\mathcal{H} \left[\prod_{a=1}^{A}\prod_{t=1}^{T} \eta_{ta}^{C_{ta}^{TA}}\right] \left[\prod_{d=1}^{D} (\frac{1}{A_d})^{N_d}\right] \\
&\quad \left[\prod_{t=1}^{T} \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \prod_{w=1}^{W} \phi_{wt}^{\beta_w - 1}\right)\right] \left[\prod_{a=1}^{A} \left(\frac{\Gamma(T\alpha)}{\Gamma(\alpha)^T} \prod_{t=1}^{T} \beta_{ta}^{\alpha_t^{(k)} - 1}\right)\right] dx d\Phi d\mathcal{H} \\
&= \int_x \int_\Phi \int_\mathcal{H} \left[\prod_{a=1}^{A}\prod_{t=1}^{T} \eta_{ta}^{C_{ta}^{TA}}\right] \left[\prod_{d=1}^{D} (\frac{1}{A_d})^{N_d}\right] \\
&\quad \left[\prod_{t=1}^{T}\prod_{w=1}^{W} \eta_{wt}^{C_{wt}^{WT} + \beta_w - 1}\right] \left[\prod_{a=1}^{A}\prod_{t=1}^{T} \phi_{ta}^{C_{ta}^{TA} + \alpha_t^{(k)} - 1}\right] dx d\Phi d\mathcal{H} \\
&= \left[\prod_{a=1}^{A}\prod_{t=1}^{T} \eta_{ta}^{C_{ta}^{TA}}\right] \int_{x_{:\frac{1}{A_d}}} \left[\prod_{d=1}^{D} (\frac{1}{A_d})^{N_d}\right] dx_{:\frac{1}{A_d}} \int_\mathcal{H} \left[\prod_{t=1}^{T}\prod_{w=1}^{W} \eta_{wt}^{C_{wt}^{WT} + \beta_w - 1}\right] d\mathcal{H} \\
&\quad \int_\Phi \left[\prod_{a=1}^{A}\prod_{t=1}^{T} \phi_{ta}^{C_{ta}^{TA} + \alpha_t^{(k)} - 1}\right] d\Phi \\
&= \left[\prod_{a=1}^{A}\prod_{t=1}^{T} \eta_{ta}^{C_{ta}^{TA}}\right] \left[\prod_{d=1}^{D} (\frac{1}{A_d})^{N_d + 1}\right] \\
&\quad \left[\prod_{a=1}^{A} \frac{\prod_{t=1}^{T} \Gamma(C_{ta}^{TA} + \alpha_t^{(k)})}{\Gamma(\sum_{t'} C_{t'a}^{TA} + T\alpha^{(k)})}\right] \left[\prod_{t=1}^{T} \frac{\prod_{w=1}^{W} \Gamma(C_{wt}^{WT} + \beta_w)}{\Gamma(\sum_{w'} C_{w't}^{WT} + W\beta)}\right]
\end{aligned}
\tag{3.5}
$$

其他

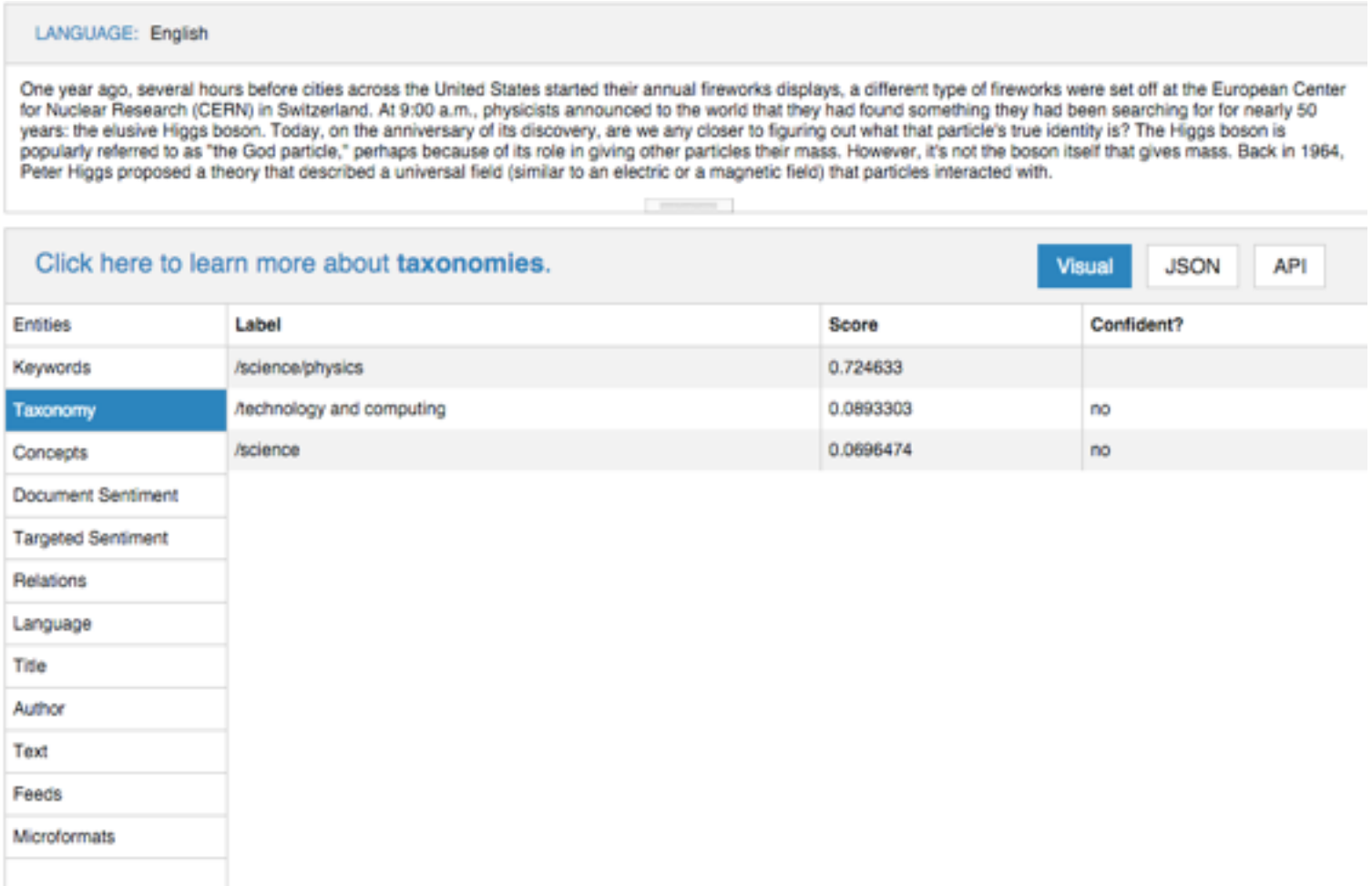
1. Alchemyapi
2. Mashape

AlchemyAPI

1. Language
2. Vision
3. News



AlchemyAPI



LANGUAGE: English

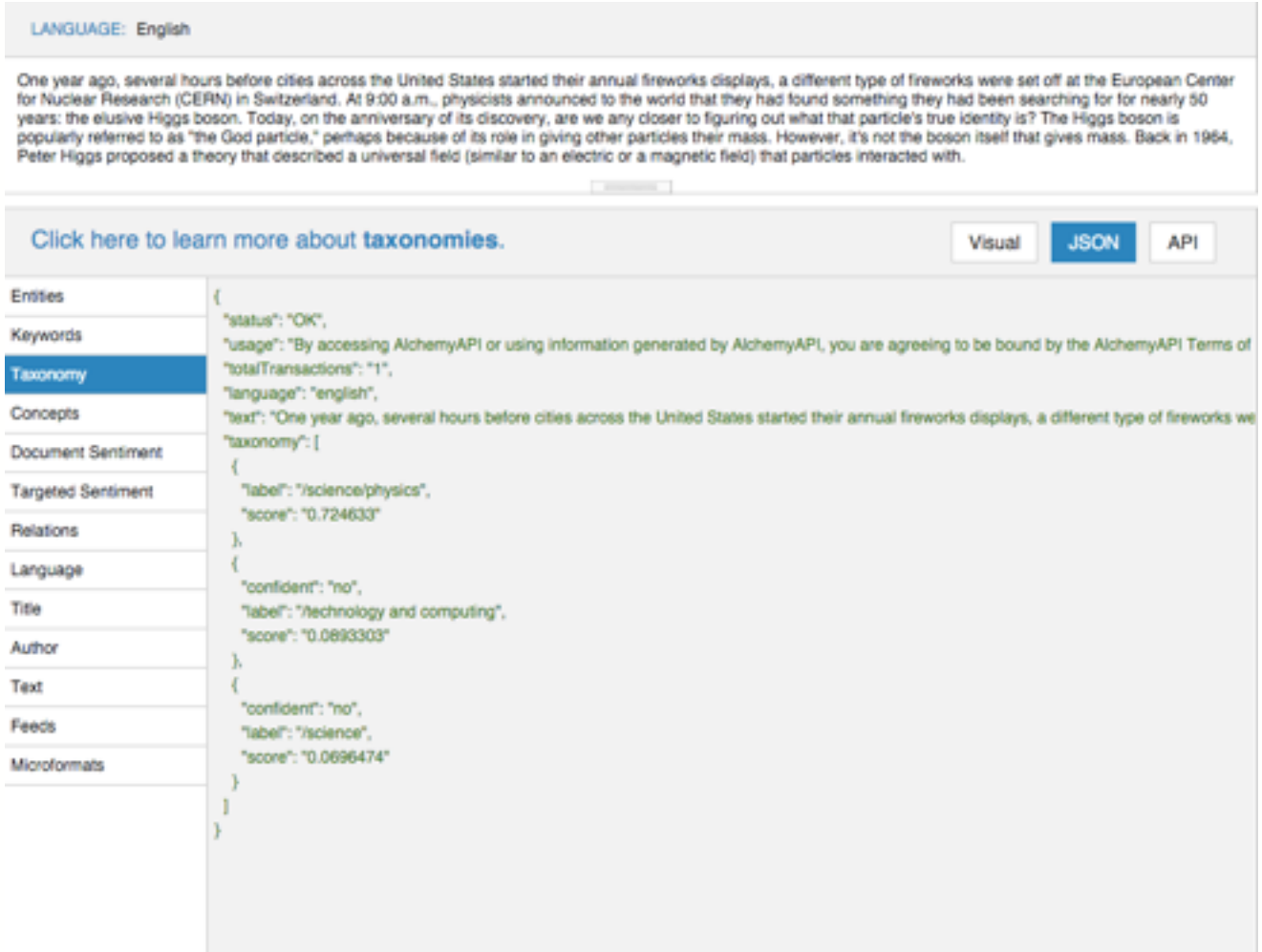
One year ago, several hours before cities across the United States started their annual fireworks displays, a different type of fireworks were set off at the European Center for Nuclear Research (CERN) in Switzerland. At 9:00 a.m., physicists announced to the world that they had found something they had been searching for for nearly 50 years: the elusive Higgs boson. Today, on the anniversary of its discovery, are we any closer to figuring out what that particle's true identity is? The Higgs boson is popularly referred to as "the God particle," perhaps because of its role in giving other particles their mass. However, it's not the boson itself that gives mass. Back in 1964, Peter Higgs proposed a theory that described a universal field (similar to an electric or a magnetic field) that particles interacted with.

Click here to learn more about **taxonomies**.

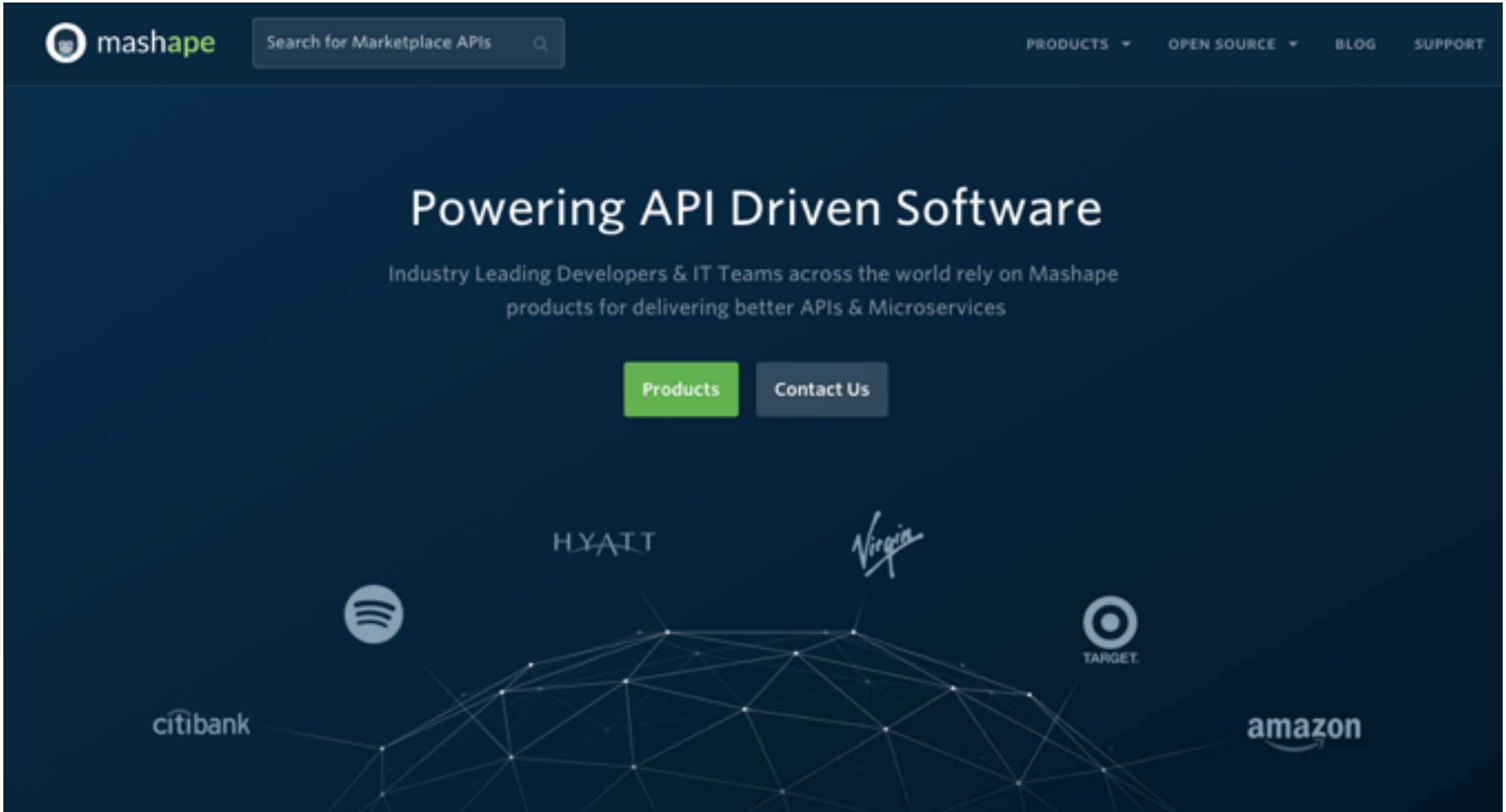
Visual | JSON | API

Entities	Label	Score	Confident?
Keywords	/science/physics	0.724633	
Taxonomy	/technology and computing	0.0893303	no
Concepts	/science	0.0696474	no
Document Sentiment			
Targeted Sentiment			
Relations			
Language			
Title			
Author			
Text			
Feeds			
Microformats			

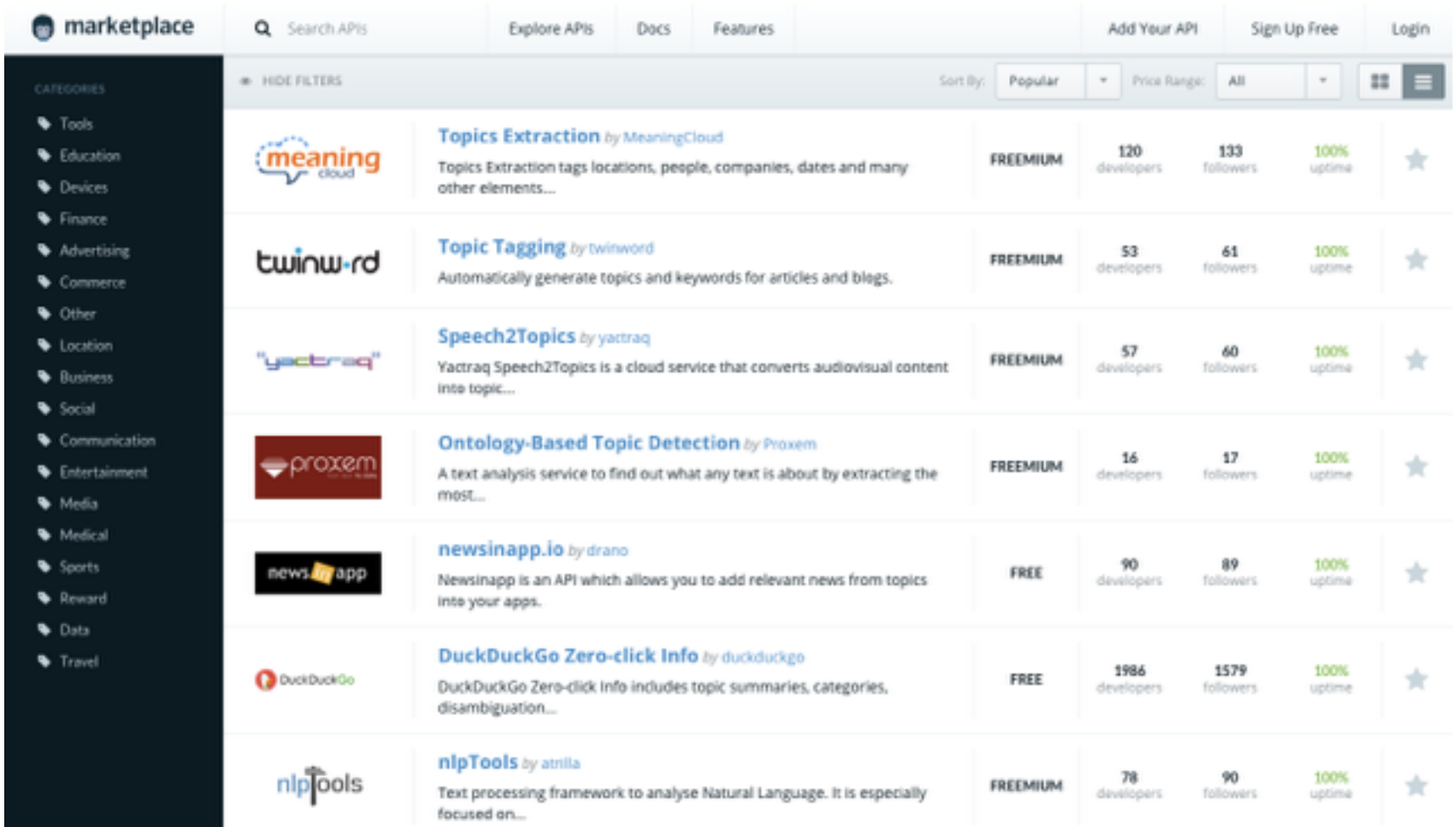
AlchemyAPI


LANGUAGE: English
One year ago, several hours before cities across the United States started their annual fireworks displays, a different type of fireworks were set off at the European Center for Nuclear Research (CERN) in Switzerland. At 9:00 a.m., physicists announced to the world that they had found something they had been searching for for nearly 50 years: the elusive Higgs boson. Today, on the anniversary of its discovery, are we any closer to figuring out what that particle's true identity is? The Higgs boson is popularly referred to as "the God particle," perhaps because of its role in giving other particles their mass. However, it's not the boson itself that gives mass. Back in 1964, Peter Higgs proposed a theory that described a universal field (similar to an electric or a magnetic field) that particles interacted with.
Click here to learn more about taxonomies.
Visual
JSON
API
Entities
Keywords
Taxonomy
Concepts
Document Sentiment
Targeted Sentiment
Relations
Language
Title
Author
Text
Feeds
Microformats
{
"status": "OK",
"usage": "By accessing AlchemyAPI or using information generated by AlchemyAPI, you are agreeing to be bound by the AlchemyAPI Terms of
"totalTransactions": "1",
"language": "english",
"text": "One year ago, several hours before cities across the United States started their annual fireworks displays, a different type of fireworks we
"taxonomy": [
{
"label": "/science/physics",
"score": "0.724633"
},
{
"confident": "no",
"label": "/technology and computing",
"score": "0.0893303"
},
{
"confident": "no",
"label": "/science",
"score": "0.0696474"
}
]
}

Mashape


mashape
Search for Marketplace APIs
PRODUCTS
OPEN SOURCE
BLOG
SUPPORT
Powering API Driven Software
Industry Leading Developers & IT Teams across the world rely on Mashape products for delivering better APIs & Microservices
Products
Contact Us
HYATT
Virgin
TARGET
citibank
amazon

Mashape


marketplace
Search APIs
Explore APIs
Docs
Features
Add Your API
Sign Up Free
Login
CATEGORIES
Tools
Education
Devices
Finance
Advertising
Commerce
Other
Location
Business
Social
Communication
Entertainment
Media
Medical
Sports
Reward
Data
Travel
HIDE FILTERS
Sort By: Popular
Price Range: All
Topics Extraction by MeaningCloud
Topics Extraction tags locations, people, companies, dates and many other elements...
FREEMIUM
120 developers
133 followers
100% uptime
Topic Tagging by twinword
Automatically generate topics and keywords for articles and blogs.
FREEMIUM
53 developers
61 followers
100% uptime
Speech2Topics by yactraq
Yactraq Speech2Topics is a cloud service that converts audiovisual content into topic...
FREEMIUM
57 developers
60 followers
100% uptime
Ontology-Based Topic Detection by Proxem
A text analysis service to find out what any text is about by extracting the most...
FREEMIUM
16 developers
17 followers
100% uptime
newsinapp.io by drano
Newsinapp is an API which allows you to add relevant news from topics into your apps.
FREE
90 developers
89 followers
100% uptime
DuckDuckGo Zero-click Info by duckduckgo
DuckDuckGo Zero-click Info includes topic summaries, categories, disambiguation...
FREE
1986 developers
1579 followers
100% uptime
nlpTools by atrilla
Text processing framework to analyse Natural Language. It is especially focused on...
FREEMIUM
78 developers
90 followers
100% uptime

任务

1. 安装sublime text3
2. HTML基本标签
   - (www.w3cschool.com)

下期预告

1. VPN, shadowsocks
2. 系统安装
3. 安全攻防