opentopic

COGNITIVE DIGITAL MARKETING

Let me show you what opentopic can do :)

Get Content Recommendations for defined target audience,

Automate your content marketing campaigns,

Increase number of ‘transactions’ for your business

History Of Opentopic

1. Blog Platform with curated content (big mess …. :)
2. Rewrite for curation platform - bootstrap / twitter engine / social publications (no time to make it perfect … startup life…)
3. Redesign Architecture to make it good for smaller clients
4. Move from our parsers to diffbot
5. Move to elasticsearch as our main db storage
6. Improve UI + add more sources + add more channels + more features (trends, scaling issues, enterprise permissioning, other crazy ideas :))
7. Continuous integration, auto scaling
8. Revolution Start - Front / Backend separation - move to django rest framework and angular
9. Architecture cleanup - salt, nagios...
10. Recommendations - cooperation with Watson

Technology that we are using

AWS - Amazon Web Services (EC2, RDS, SQS, KINESIS, LAMBDA, ELASTICACHE)

ElasticSearch

Diffbot

Google Translate API

IBM Watson - News API, Taxonomy API

NLTK, SCIPY

Python

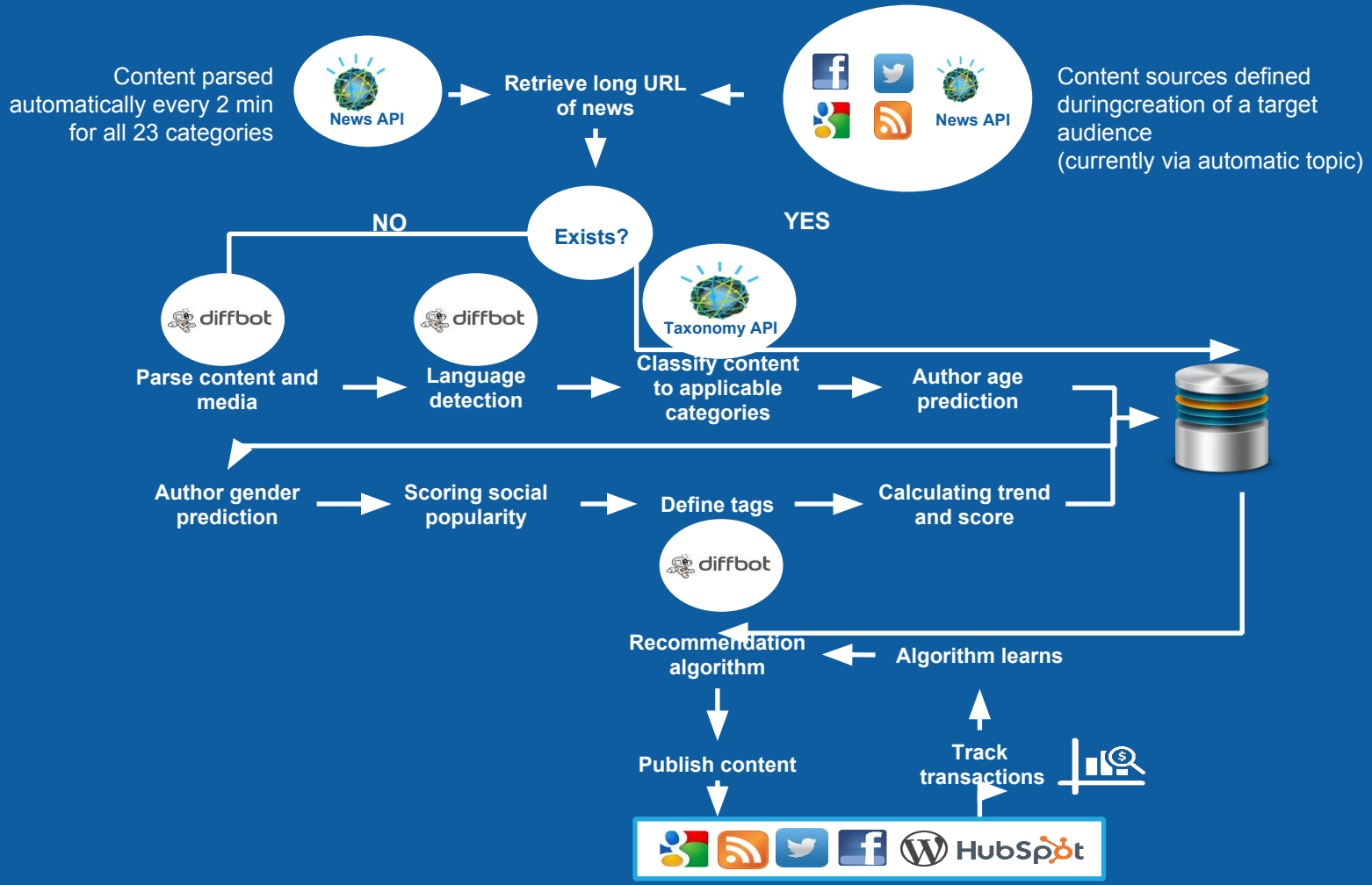
Django

Salt

Jenkins

Nagios

Slack


Content parsed automatically every 2 min for all 23 categories
News API
Retrieve long URL of news
News API
Content sources defined duringcreation of a target audience (currently via automatic topic)
NO
Exists?
YES
diffbot
Parse content and media
diffbot
Language detection
Taxonomy API
Classify content to applicable categories
Author age prediction
Author gender prediction
Scoring social popularity
Define tags
Calculating trend and score
diffbot
Recommendation algorithm
Algorithm learns
Publish content
Track transactions
HubSpot

Numbers

20 mln news (last 3 months)

25 ec2 machines (75 vCPU, 130 GB Memory)

321 071 623 tasks

2 672 404 s3 requests

4 527 GB of data

50 000 sources + 250 000 sources from IBM

Our client's results speak for themselves

"We increased our newsletter click through rate by 350% thanks to the Opentopic tool"

pitney bowes

"Our number of followers increased by 10% in the first 20 days"

CLOSERS
AV SÄLJARE FÖR SÄLJARE

"Our blog visits increased almost 100% in the first month of using Opentopic."

LOAD IMPACT

"Within a few weeks of using Opentopic, our website traffic had gone up by 190%"

VINCI

"Over a 23 week period, content published through Opentopic generated an average increase in website traffic of 180%"

opentopic

Our client's results speak For themselves

"We use Opentopic to animate our social channels and have seen an increase of 110% in traffic"

"Since using Opentopic we get an average of 2 retweets per post"

Content categorization - Naive Bayes classifier

Probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions between the features

Learning from supervision

Maximum likelihood

Finds application in automatic medical diagnosis

Python implementation:

- nltk (Natural Language Toolkit)

Content categorization - Natural Language Processing

Natural Language Processing:

- a field of computer science, artificial intelligence, and computational linguistics
- used within area of human–computer interaction
- summarizations
- translations
- named entity recognition
- OCR
- Question answering
- Speech recognition
- Speech processing

Content categorization - nltk

NLTK:

- eliminating stop words (the, is, at, which, and on)
- grouping words into phrases
- classifiers:
  - ConditionalExponentialClassifier
  - DecisionTreeClassifier
  - MaxentClassifier
  - NaiveBayesClassifier
  - WekaClassifier

```
nltk.corpus.stopwords.words('en')
top_words = nltk.FreqDist(words).most_common()
# 1000
features= fn(top_words)  # assign words to cat.
b = nltk.NaiveBayesClassifier.train(features)

top_words = b.top_words
features = Content(words).features
(top_words=top_words)
category = b.classify(features)
```

IBM Watson

IBM Watson:

- a question answering computer system capable of answering questions posed in natural language
- access to 200 million pages of structured and unstructured content consuming four terabytes of disk storage[8] including the full text of Wikipedia
- 750 servers, 3.5 GHz POWER7 eight core processor, with four threads per core: 2,880 processor threads and 16 terabytes of RAM
- an process 500 gigabytes, the equivalent of a million books, per second
- 1 mln $ prize - Jeopardy!

What’s next?

What’s next?

- SVM (SVC)
- numpy, scipy
- ...

opentopic

GET IN TOUCH

Cognitive Digital Marketing

291 Broadway, Suite 302
New York, NY 10007

Lukasz Prusiel
Lead Developer Opentopic
luke@opentopic.com

291 Broadway, Suite 302
New York, NY 10007

Tomasz Roszko
CTO Opentopic, CEO TJ.Software
tom@opentopic.com