COMP9321: Data services engineering

Semester 1, 2018,

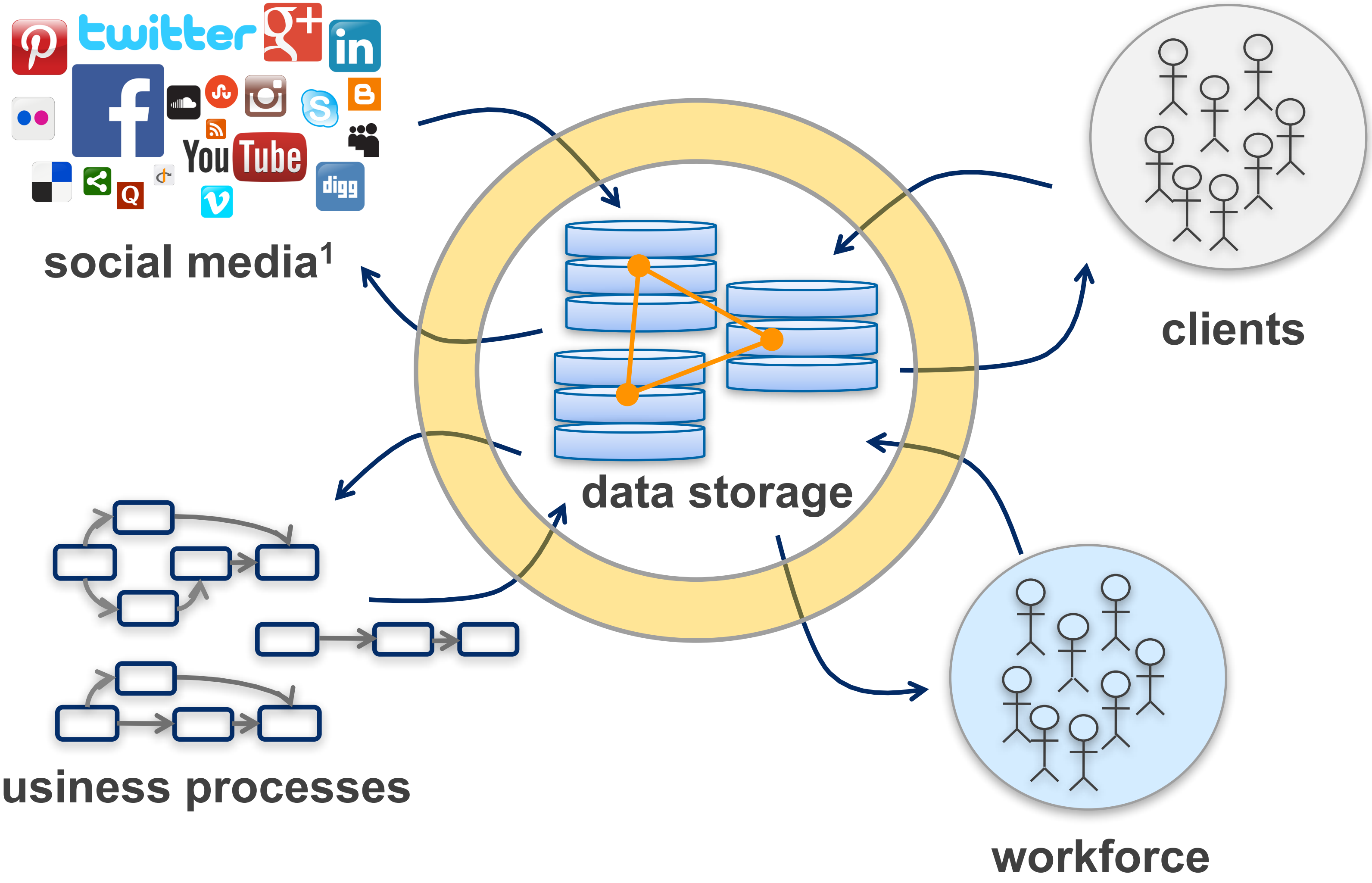
Week 9 Data analytics services: An overview

By Carlos Rodríguez, CSE UNSW

Outline

- Data driven organizations and data analytics
- Data analytics technologies and services
- Example 1: NLP services
- Example 2: Machine learning services

Data-driven organizations



social media[1]

clients

data storage

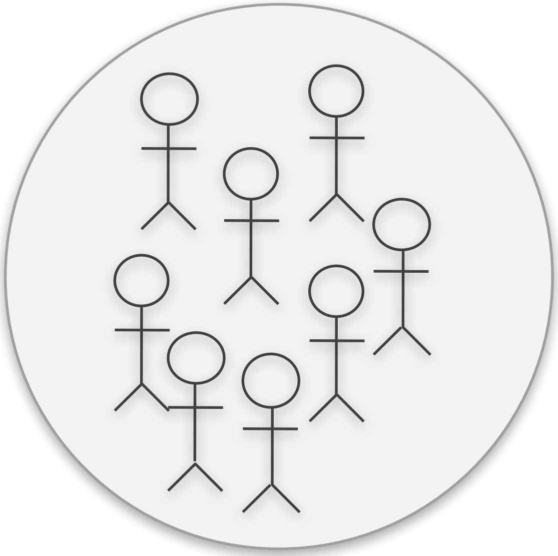
business processes

workforce

image source: [1]commons.wikimedia.org

Data-driven organizations and data analytics

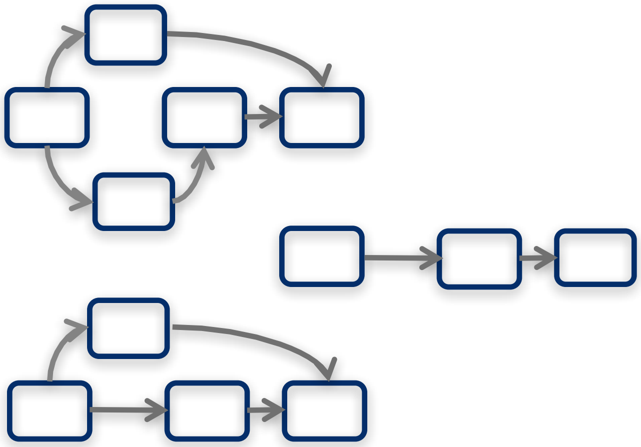
- Product and service recommendation
- Customer support
- Dashboard and reporting services
- Customer engagement
- Promotions and deals
- Product and service customization
- Communication



Clients

Data-driven organizations and data analytics

- Key process performance indicators
- Process execution predictions
- Decision making support services
- Process mining
- Dynamic process adaptation
- People to task assignment
- Compliance verification



business processes

Data-driven organizations and data analytics

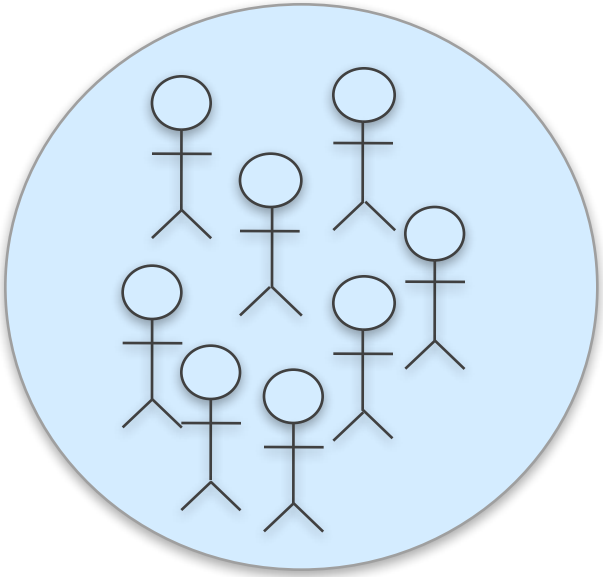
- Product and service advertisement
- Sentiment analysis
- Demographics analysis
- Virality
- Social network insights

social media[1]

image source: [1]commons.wikimedia.org

Data-driven organizations and data analytics

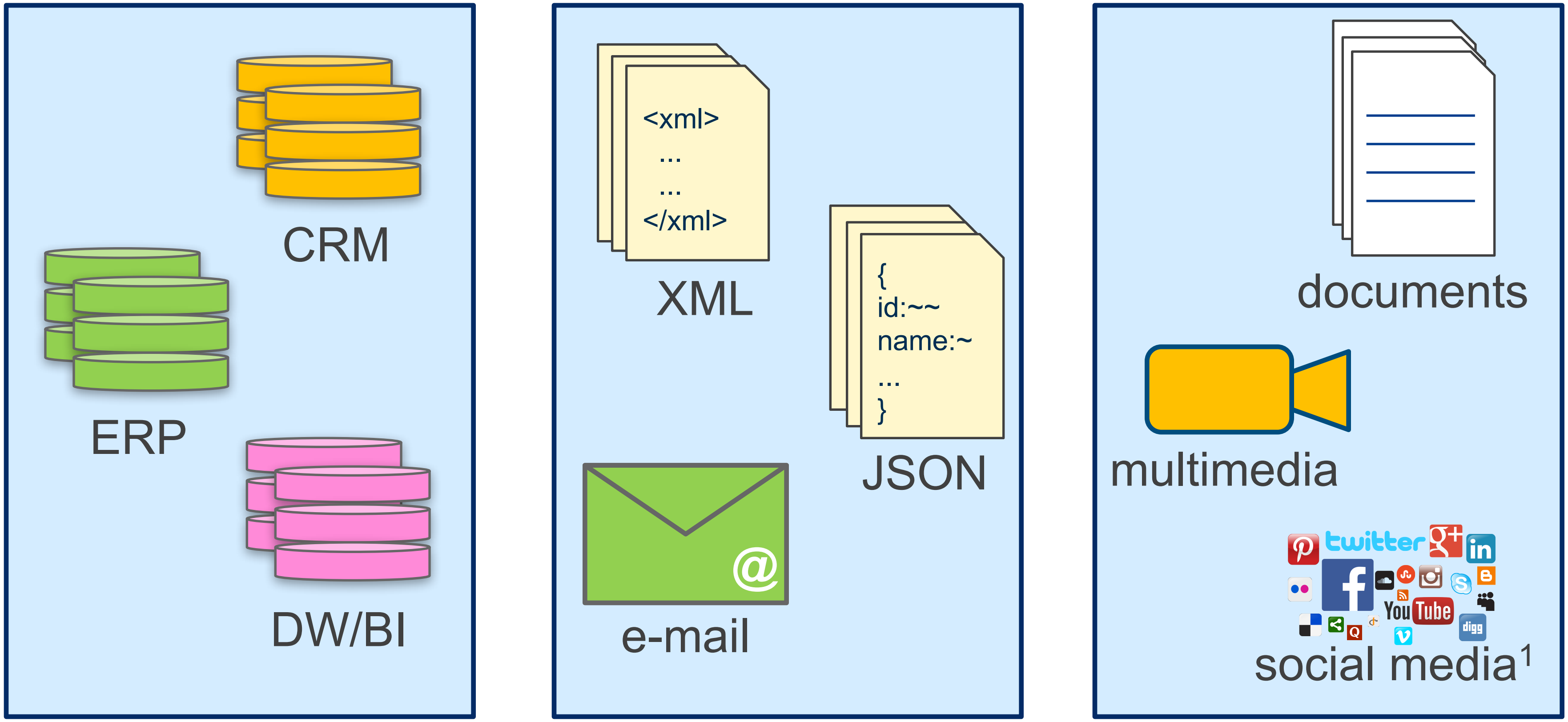
- Productivity
- Work planning
- Employee engagement
- Recruitment
- Training
- Job satisfaction
- Support



workforce

image source: [1]commons.wikimedia.org

Data used for analytics purposes



CRM

ERP

DW/BI

```
<xml>
  …
  …
</xml>
```

XML

```
{
id:~~
name:~
…
}
```

JSON

e-mail

documents

multimedia

social media[1]

structured data

semi-structured data

unstructured data

image source: [1]commons.wikimedia.org

Data analytics architecture

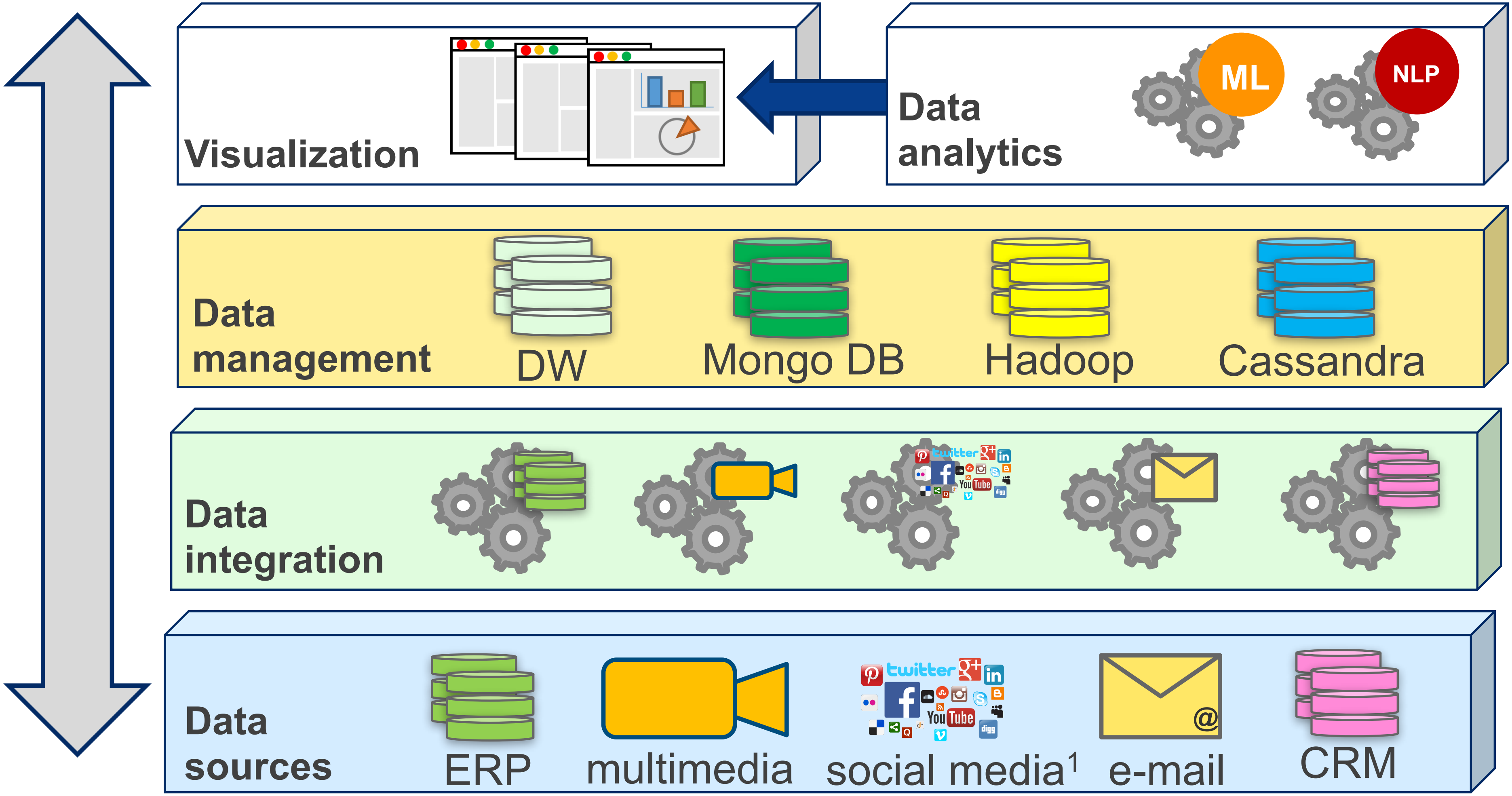


image source: [1]commons.wikimedia.org

Data analytics technologies and services

Marketing analytics

- Data collection and management
- Overall performance of your products (websites, apps, etc.)
- Evaluate the content in your products
- Understand your audience
- Data visualization and analytics

image source: commons.wikimedia.org

Marketing analytics

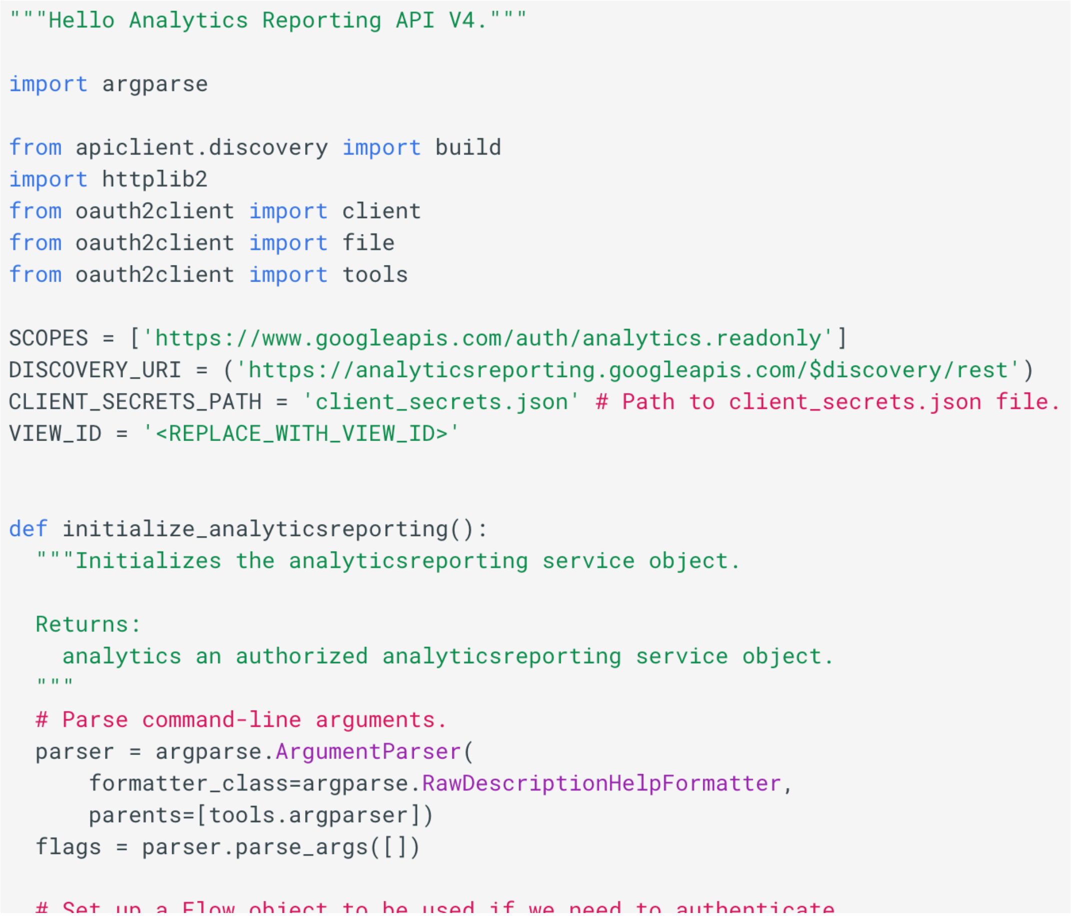
Data access services (APIs)

- Reporting API
- Real time API
- Multi-channel Funnels API

image source: commons.wikimedia.org

Marketing analytics

Analytics reporting API v4



```python
"""Hello Analytics Reporting API V4."""

import argparse

from apiclient.discovery import build
import httplib2
from oauth2client import client
from oauth2client import file
from oauth2client import tools

SCOPES = ['https://www.googleapis.com/auth/analytics.readonly']
DISCOVERY_URI = ('https://analyticsreporting.googleapis.com/$discovery/rest')
CLIENT_SECRETS_PATH = 'client_secrets.json' # Path to client_secrets.json file.
VIEW_ID = '<REPLACE_WITH_VIEW_ID>'


def initialize_analyticsreporting():
  """Initializes the analyticsreporting service object.

  Returns:
    analytics an authorized analyticsreporting service object.
  """
  # Parse command-line arguments.
  parser = argparse.ArgumentParser(
      formatter_class=argparse.RawDescriptionHelpFormatter,
      parents=[tools.argparser])
  flags = parser.parse_args([])

  # Set up a Flow object to be used if we need to authenticate
```

image sources: www.google.com/analytics, commons.wikimedia.org

Machine Learning on AWS

- Provides Machine Learning as a service
- Allows for the integration of ML APIs with other tools in the platform, including data lakes and database tools
- Provides support for major ML frameworks like TensorFlow, PyTorch and Caffe2
- Access to computing infrastructure (e.g., GPUs and FPGAs)

ML on AWS[1]

image source: wikipedia.org

Machine Learning on AWS

Vision services:

Services for identifying objects, people, text and scenes from images and videos.

Conversational bots

Services for building conversational bots using voice and text, with functionalities for speech recognition and natural language understanding.

Language services

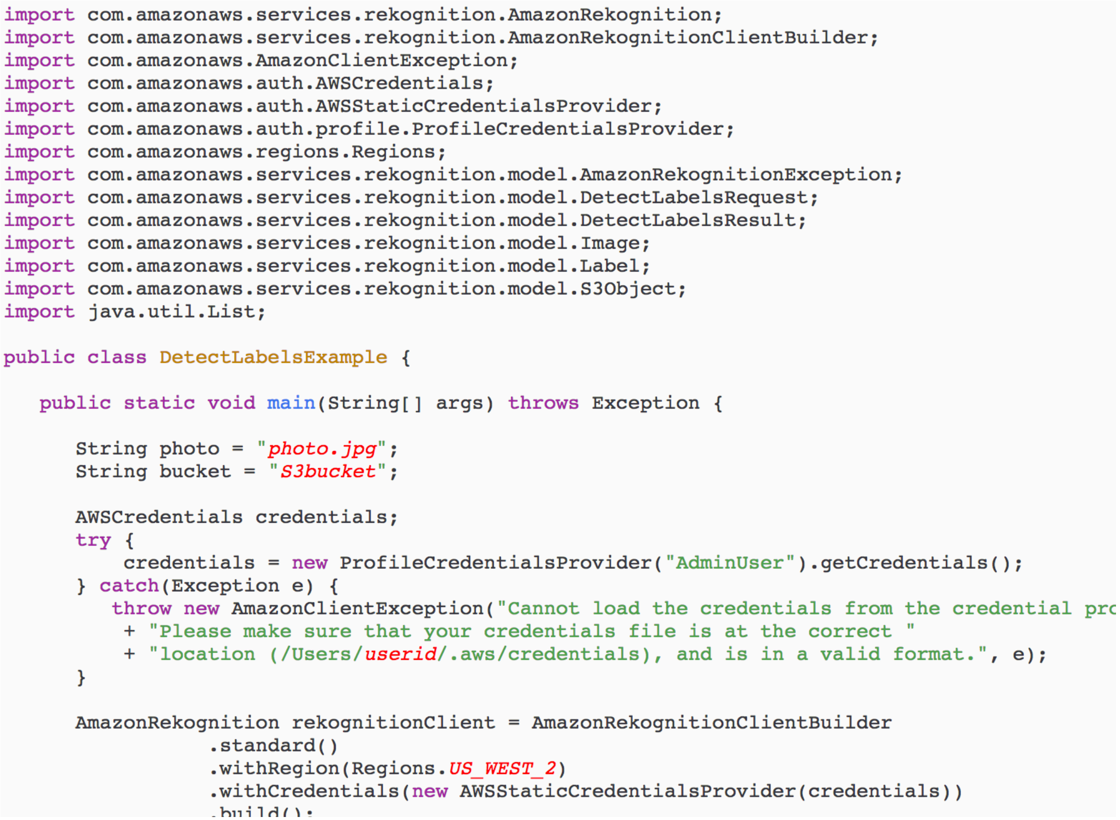
Services for translating, transcribing and speech synthesis.

ML on AWS[1]

image source: wikipedia.org

Machine Learning on AWS

Label detection API



```
import com.amazonaws.services.rekognition.AmazonRekognition;
import com.amazonaws.services.rekognition.AmazonRekognitionClientBuilder;
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.rekognition.model.AmazonRekognitionException;
import com.amazonaws.services.rekognition.model.DetectLabelsRequest;
import com.amazonaws.services.rekognition.model.DetectLabelsResult;
import com.amazonaws.services.rekognition.model.Image;
import com.amazonaws.services.rekognition.model.Label;
import com.amazonaws.services.rekognition.model.S3Object;
import java.util.List;

public class DetectLabelsExample {

   public static void main(String[] args) throws Exception {

      String photo = "photo.jpg";
      String bucket = "S3bucket";

      AWSCredentials credentials;
      try {
          credentials = new ProfileCredentialsProvider("AdminUser").getCredentials();
      } catch(Exception e) {
         throw new AmazonClientException("Cannot load the credentials from the credential pro
          + "Please make sure that your credentials file is at the correct "
          + "location (/Users/userid/.aws/credentials), and is in a valid format.", e);
      }

      AmazonRekognition rekognitionClient = AmazonRekognitionClientBuilder
                 .standard()
                 .withRegion(Regions.US_WEST_2)
                 .withCredentials(new AWSStaticCredentialsProvider(credentials))
                 .build();
```

ML on AWS[1]

image sources: wikipedia.org, aws.amazon.com/documentation/rekognition/

Natural language conversational services

- Allows for building text and voice conversational interfaces
- Provides integration with other products such as Facebook Messenger, Amazon Alexa other platforms
- Leverages on domain knowledge and natural language understanding to comprehend what the end-user is saying

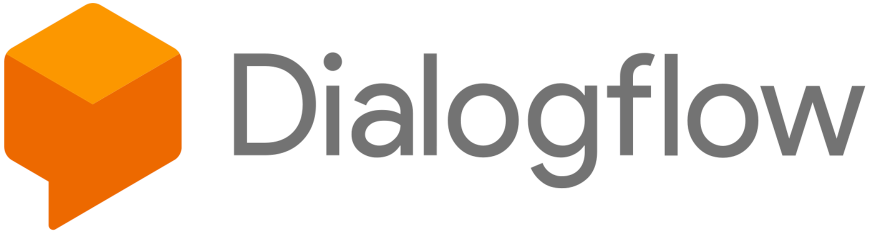


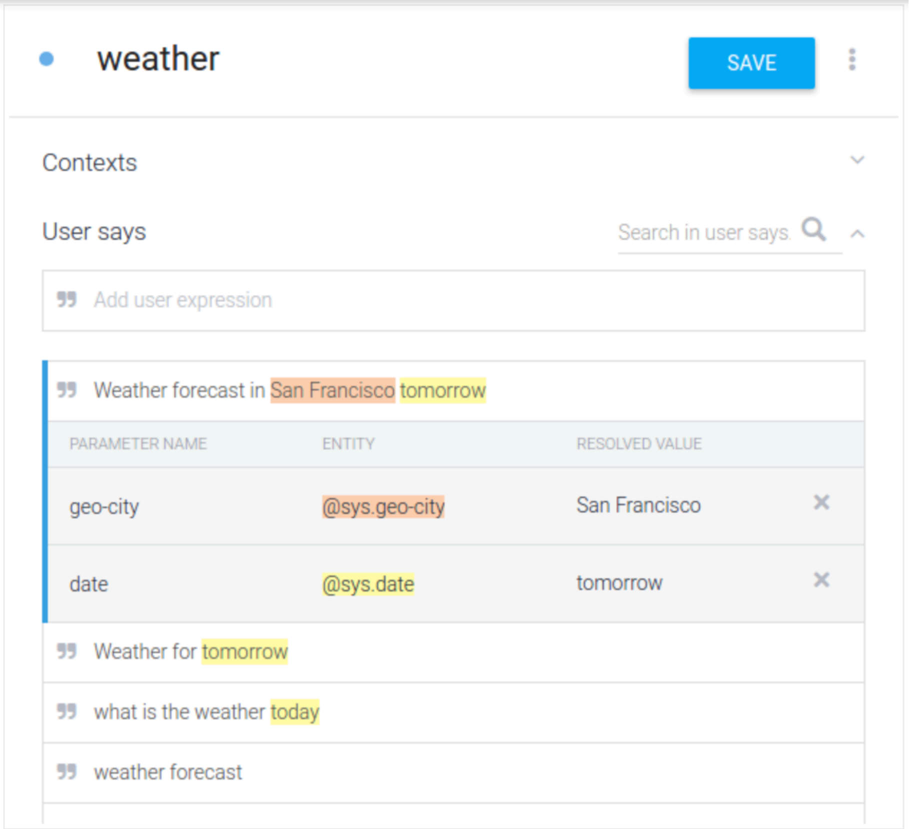
image source: wikipedia.org

Natural language conversational services

- Captures different **expressions** user can use in a conversation
- Helps in mapping such user expressions to concrete **Intents.**
- Helps in mapping **entities** found in expressions to **parameters** needed in intents
- Intents are then associated to **actions**, which are essentially API calls
- Builds the conversational interface

image source: wikipedia.org

Natural language conversational services



weather

SAVE

Contexts

User says

Search in user says

Add user expression

Weather forecast in San Francisco tomorrow

PARAMETER NAME	ENTITY	RESOLVED VALUE
geo-city	@sys.geo-city	San Francisco
date	@sys.date	tomorrow

Weather for tomorrow

what is the weather today

weather forecast

Dialogflow

image sources: wikipedia.org, dialogflow.com

Example 1: NLP services

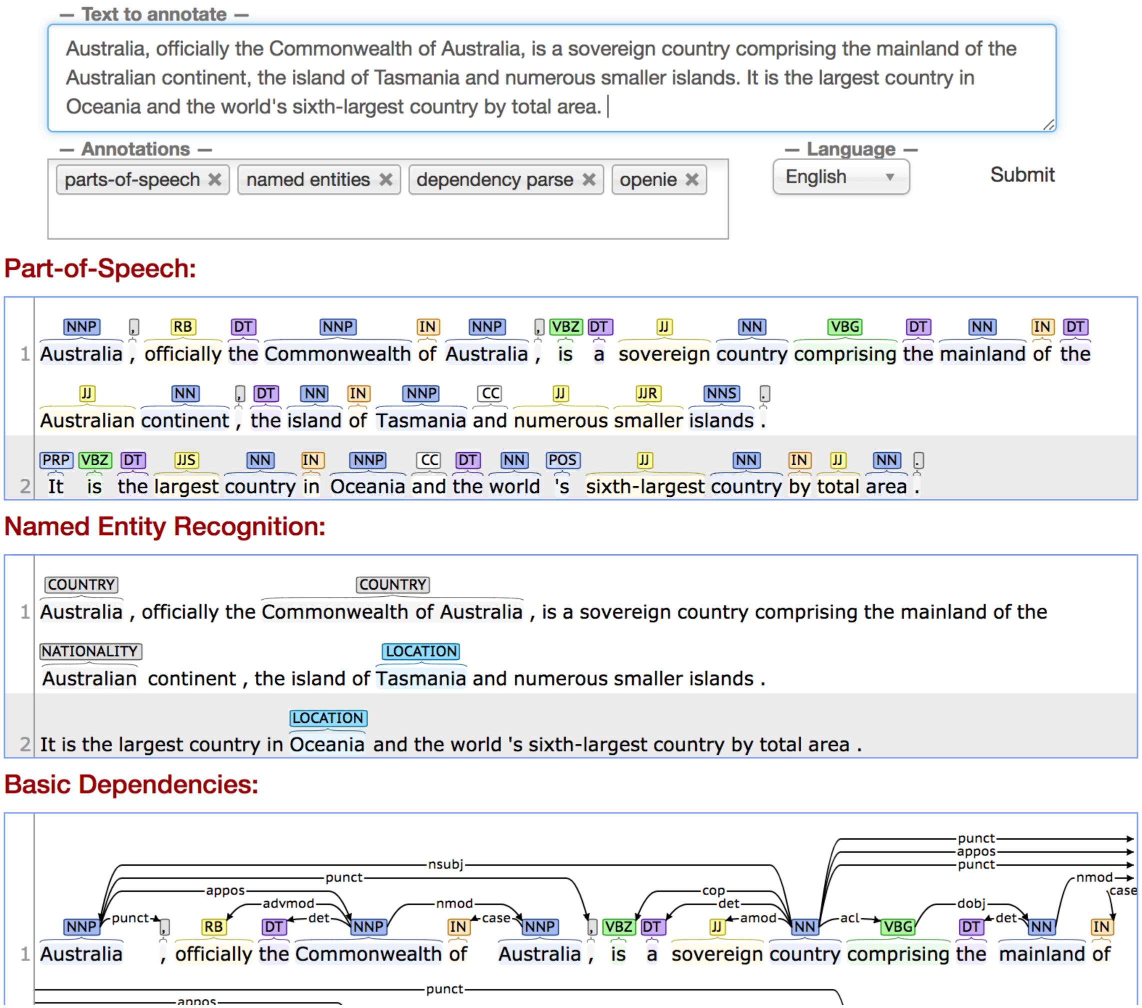
Stanford Core NLP

Natural language functionalities

- Part of speech
- Lemmas
- Named entities
- Dependency parse
- Relation extraction
- Coreference
- Sentiment analysis

Stanford CoreNLP

Stanford Core NLP



— Text to annotate —

Australia, officially the Commonwealth of Australia, is a sovereign country comprising the mainland of the Australian continent, the island of Tasmania and numerous smaller islands. It is the largest country in Oceania and the world's sixth-largest country by total area.

— Annotations —

parts-of-speech ✕ named entities ✕ dependency parse ✕ openie ✕

— Language —

English ▾

Submit

Part-of-Speech:

1 Australia , officially the Commonwealth of Australia , is a sovereign country comprising the mainland of the Australian continent , the island of Tasmania and numerous smaller islands .

2 It is the largest country in Oceania and the world 's sixth-largest country by total area .

NNP , RB DT NNP IN NNP , VBZ DT JJ NN VBG DT NN IN DT
JJ NN , DT NN IN NNP CC JJ JJR NNS .
PRP VBZ DT JJS NN IN NNP CC DT NN POS JJ NN IN JJ NN .

Named Entity Recognition:

1 Australia , officially the Commonwealth of Australia , is a sovereign country comprising the mainland of the Australian continent , the island of Tasmania and numerous smaller islands .

2 It is the largest country in Oceania and the world 's sixth-largest country by total area .

Basic Dependencies:

1 Australia , officially the Commonwealth of Australia , is a sovereign country comprising the mainland of

Stanford CoreNLP

image and text sources: corenlp.run, wikipedia.org,

Stanford Core NLP Server: Installation

Server that provides a Web API to access the main functionalities provided by Stanford Core NLP

Stanford Core NLP (and server) can be downloaded from:

https://goo.gl/fBaoBL

The server is implemented in Java, and its APIs (libraries) can also be used inside your code.

image and text sources: corenlp.run, wikipedia.org,

Stanford Core NLP Server: Installation

1) Download the server: stanford-corenlp-full-2017-06-09.zip

   https://goo.gl/fBaoBL

2) Unzip the file and open the created directory.

3) Run the server:

```
$ java -mx4g -cp "*" edu.stanford.nlp.pipeline.StanfordCoreNLPServer -port 9000
```

image and text sources: corenlp.run, wikipedia.org,

Stanford Core NLP Server: Services

Annotation service

End point: POST /?properties={...}

Query TokensRegex service

End point: POST /tokensregex?pattern={..}&filter={..}

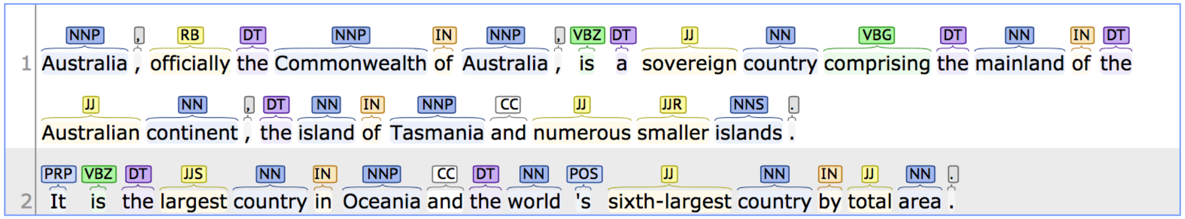
Query Semregex service

End point: POST /semgrex?pattern={..}&filter={..}

image and text sources: corenlp.run, wikipedia.org,

Stanford Core NLP Server: Annotation services

Part-of-Speech:



1 Australia , officially the Commonwealth of Australia , is a sovereign country comprising the mainland of the Australian continent , the island of Tasmania and numerous smaller islands .

2 It is the largest country in Oceania and the world 's sixth-largest country by total area .

NNP , RB DT NNP IN NNP , VBZ DT JJ NN VBG DT NN IN DT
JJ NN , DT NN IN NNP CC JJ JJR NNS .
PRP VBZ DT JJS NN IN NNP CC DT NN POS JJ NN IN JJ NN .

End point: POST /?properties={“annotators”:“pos”, “outputFormat”:“json”}

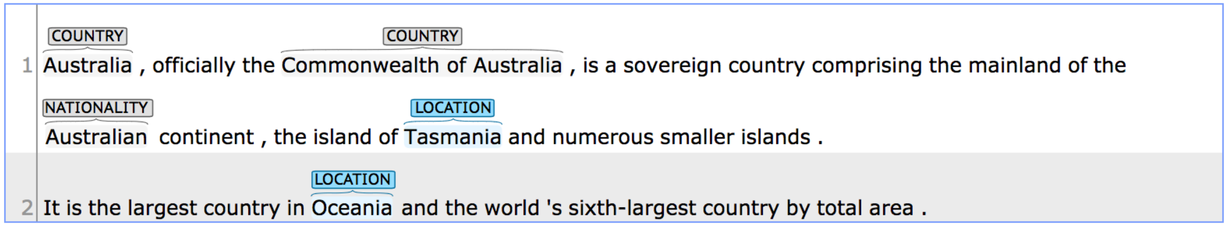
```
$ wget --post-data 'Australia, officially the Commonwealth of Australia, is a sovereign
country comprising the mainland of the Australian continent.'
'localhost:9000/?properties={"annotators":"pos","outputFormat":"json"}' -O -
```

```
{
   "sentences": [{"index":0,"tokens": [{"index":1,"word":"Australia",
   "originalText":"Australia","characterOffsetBegin":0,"characterOffsetEnd":9,"pos":
   "NNP","before":"","after":""},{"index":2,"word":",","originalText":",",
   "characterOffsetBegin":9,"characterOffsetEnd":10,"pos":",","before":"","after":" "},
  {"index":3,"word":"officially","originalText":"officially","characterOffsetBegin":11,
   "characterOffsetEnd":21 ...
}
```

image and text sources: corenlp.run, wikipedia.org,

Stanford Core NLP Server: Annotation services

Named Entity Recognition:



1 Australia , officially the Commonwealth of Australia , is a sovereign country comprising the mainland of the Australian continent , the island of Tasmania and numerous smaller islands .

2 It is the largest country in Oceania and the world 's sixth-largest country by total area .

End point: POST /?properties={“annotators”:“ner”, “outputFormat”:“json”}

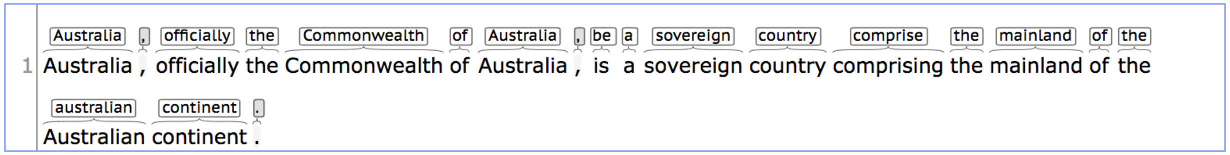
```
$ wget --post-data 'Australia, officially the Commonwealth of Australia, is a sovereign
country comprising the mainland of the Australian continent.'
'localhost:9000/?properties={"annotators":"ner","outputFormat":"json"}' -O -
```

```
{
  "sentences":[{"index":0,"tokens":[{"index":1,"word":"Australia","originalText":
  "Australia","lemma":"Australia","characterOffsetBegin":0,"characterOffsetEnd":9,
  "pos":"NNP","ner":"LOCATION","before":"","after":""},{"index":2,"word":",",
  "originalText":",","lemma":",","characterOffsetBegin":9,"characterOffsetEnd":10,
  "pos":",","ner":"O","before":"","after":" "}, ...
}
```

image and text sources: corenlp.run, wikipedia.org,

Stanford Core NLP Server: Annotation services

Lemmas:



```
  Australia , officially the Commonwealth of Australia , be a sovereign country comprise the mainland of the
1 Australia , officially the Commonwealth of Australia , is a sovereign country comprising the mainland of the
  australian continent .
  Australian continent .
```

End point: POST /?properties={“annotators”:“lemma”, “outputFormat”:“json”}

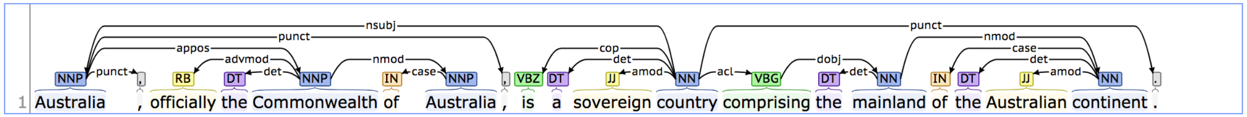
```
$ wget --post-data 'Australia, officially the Commonwealth of Australia, is a sovereign country comprising the mainland of the Australian continent.' 'localhost:9000/?properties={"annotators":"lemma","outputFormat":"json"}' -O -
```

```
{
  ... {"index":13,"word":"comprising","originalText":"comprising","lemma":
  "comprise","characterOffsetBegin":76,"characterOffsetEnd":86,"pos":"VBG",
  "before":" ","after":" "},{"index":14,"word":"the","originalText":"the","lemma":
  "the", "characterOffsetBegin":87,"characterOffsetEnd":90,"pos":"DT",
  "before":"  ","after":"  "}...
}
```

image and text sources: corenlp.run, wikipedia.org,

Stanford Core NLP Server: Annotation services

Basic Dependencies:


nsubj
punct
appos
advmod
det
punct
nmod
case
cop
det
amod
acl
punct
nmod
case
det
dobj
det
amod
NNP
,
RB
DT
NNP
IN
NNP
,
VBZ
DT
JJ
NN
VBG
DT
NN
IN
DT
JJ
NN
.
1
Australia , officially the Commonwealth of Australia , is a sovereign country comprising the mainland of the Australian continent .

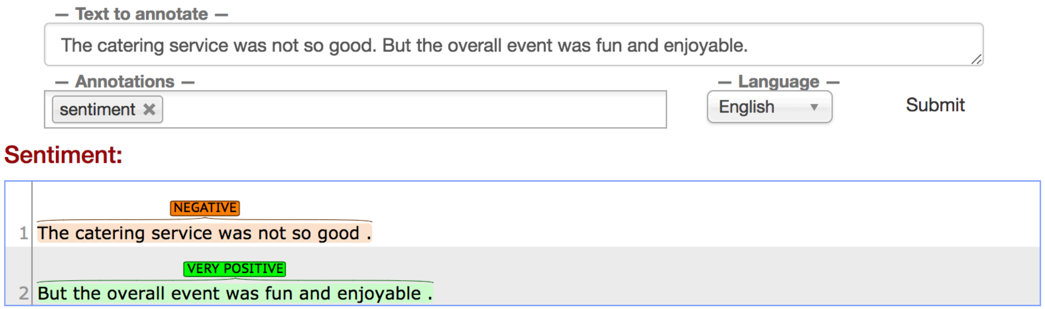
End point: POST /?properties={“annotators”:“depparse”, “outputFormat”:“json”}

```
$ wget --post-data 'Australia, officially the Commonwealth of Australia, is a sovereign
country comprising the mainland of the Australian continent.'
'localhost:9000/?properties={"annotators":"depparse","outputFormat":"json"}' -O -
```

```
{
  "sentences":[{"index":0,"basicDependencies":[{"dep":"ROOT","governor":0,
  "governorGloss":"ROOT","dependent":12,"dependentGloss":"country"},
 {"dep":"nsubj","governor":12,"governorGloss":"country","dependent":1,
  "dependentGloss":"Australia"},{"dep":"punct","governor":1,"governorGloss":
  "Australia","dependent":2,"dependentGloss":","},{"dep":"advmod","governor":5,
  "governorGloss":"Commonwealth","dependent":3,"dependentGloss":"officially"} ...
}
```

image and text sources: corenlp.run, wikipedia.org,

Stanford Core NLP Server: Annotation services



— Text to annotate —

The catering service was not so good. But the overall event was fun and enjoyable.

— Annotations —

sentiment ✖

— Language —

English

Submit

Sentiment:

NEGATIVE

1 The catering service was not so good .

VERY POSITIVE

2 But the overall event was fun and enjoyable .

End point: POST /?properties={“annotators”:“sentiment”, “outputFormat”:“json”}

```
$ wget --post-data 'The catering service was not good. But the overall event was fun and enjoyable.' 'localhost:9000/?properties={"annotators":"sentiment","outputFormat":"json"}' -O -
```

```
{
  "sentiment":"Negative","tokens":[{"index":1,"word":"The","originalText":"The",
  "characterOffsetBegin":0,  ... ","after":" "},{"index":5,"word":"not","originalText":
  "not","characterOffsetBegin":25,"characterOffsetEnd":28,"pos":"RB",
  "before":"  ","after":" "},{"index":6,"word":"good","originalText":"good",
  "characterOffsetBegin":29
}
```

Image source: corenlp.run

Example 2: Machine Learning services

Machine learning services

- Open source Machine Learning server
- Build and deploy ML web services
- Customizable, template-based engines
- Real-time ML services both for training and prediction
- Built on top of state of the art algorithms and tools

PredictionIO

source: prediction.io

Machine learning services

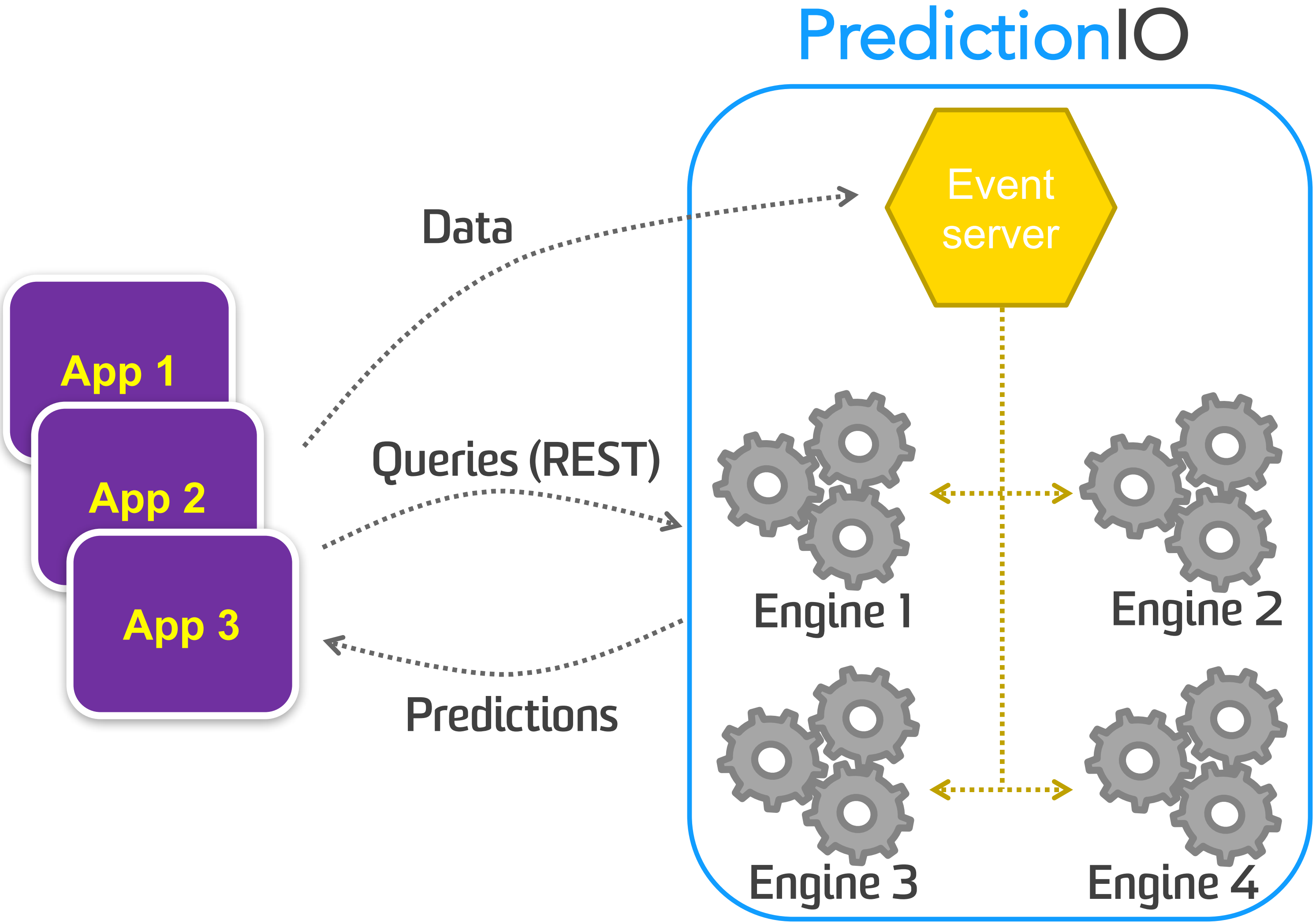
Engine templates

- Recommenders
- Classification
- Regression
- NLP
- Clustering
- Similarity

PredictionIO

source: prediction.io

PredictionIO: Overall architecture



source: prediction.io

PredictionIO: Server installation

- You can install PredictionIO in one of the following ways:
  - Installation from source code
  - Installation with Docker

- We follow the second option (Installation with Docker)
  1. Install Docker (https://www.docker.com/)
  2. Use the following container (and installation instructions):
     - https://github.com/steveny2k/docker-predictionio
  3. Run PredictionIO:

```
$ docker run -it -p 8000:8000 steveny/predictionio:0.12.0 /bin/bash
```

source: prediction.io

PredictionIO: Starting services

- Start all necessary services, including HBase, Elastic Search and PredictionIO:

```
$ pio-start-all
```

- Check if all services are up and running

```
$ jps –l

144 org.apache.hadoop.hbase.master.HMaster

576 sun.tools.jps.Jps

371 org.apache.predictionio.tools.console.Console

58 org.elasticsearch.bootstrap.Elasticsearch
```

source: prediction.io

PredictionIO: The similar product engine

- This is an engine template that recommends products that are similar to a given product.
  - http://predictionio.apache.org/templates/similarproduct/quickstart/

1. Clone the engine template:

```
$ git clone https://github.com/apache/predictionio-template-similar-product.git MySimilarProduct

$ cd MySimilarProduct
```

2. Generate an App ID and access key

```
$ pio app new MyApp1

...

[INFO] [App$] Access Key: 3mZWDzci2D5YsqAnqNnXH9SB6R
```

source: prediction.io

PredictionIO: The similar product engine

3. Save the access key in a shell variable

```
$ export ACCESS_KEY=3mZWDzci2D5YsqAnqNnXH9SB6R
```

4. Create a new user

```
$ curl -i -X POST \
http://localhost:7070/events.json?accessKey=$ACCESS_KEY \

-H "Content-Type: application/json" \

-d '{  "event" : "$set",  "entityType" : "user",  "entityId" : "u0",  \

"eventTime" : "2014-11-02T09:39:45.618-08:00"}'
```

source: prediction.io

PredictionIO: The similar product engine

4. Create a new item

```
$ curl -i -X POST  \
http://localhost:7070/events.json?accessKey=$ACCESS_KEY \
-H "Content-Type: application/json" \
-d '{  "event" : "$set",  "entityType" : "item",  "entityId" : "i0", \
 "properties" : {    "categories" : ["c1", "c2"]  }  "eventTime" : \
 "2014-11-02T09:39:45.618-08:00"}'
```

source: prediction.io

PredictionIO: The similar product engine

5. Create a new view of item “i0” by user “u0”

```
$ curl -i -X POST \
http://localhost:7070/events.json?accessKey=$ACCESS_KEY \
-H "Content-Type: application/json" \-d '{  "event" : "view",  \
"entityType" : "user",  "entityId" : "u0",  "targetEntityType" : "item",\
"targetEntityId" : "i0",  "eventTime" : \
"2014-11-10T12:34:56.123-08:00"}'
```

6. Query all posted events

```
$ curl -i -X GET \
"http://localhost:7070/events.json?accessKey=$ACCESS_KEY"
```

source: prediction.io

PredictionIO: The similar product engine

7. Import more data (users, items and views)

```
$ easy_install predictionio
$ cd MySimilarProduct
$ python data/import_eventserver.py --access_key $ACCESS_KEY
```

8. Build the engine:

8.1 Edit file engine.json in MySimilarProduct, add “MyApp1” to appName

```
″datasource″: {
    ″params″ : {
        ″appName": "MyApp1"
    }
  },   ...
```

source: prediction.io

PredictionIO: The similar product engine

8.2 Build the engine

```
$ pio build --verbose
```

9. Train the model:

```
$ pio train
```

10. Deploy the engine

```
$ pio deploy
```

source: prediction.io

PredictionIO: The similar product engine

11. Using the engine

11.1 Get 4 items similar to item “i1”

```
$ curl -H "Content-Type: application/json" \-d '{ "items": ["i1"], "num": 4 }'  \
http://localhost:8000/queries.json
```

11.2 Get items similar to a list of items

```
$ curl -H "Content-Type: application/json" \-d '{ "items": ["i1", "i3"], "num": 10}' \
http://localhost:8000/queries.json
```

11.3 Get items in a list of categories

```
$ curl -H "Content-Type: application/json" \-d '{  "items": ["i1", "i3"],  "num": 10,  \
"categories" : ["c4", "c3"]}' \http://localhost:8000/queries.json
```

source: prediction.io

Appendix A

Using corenlp.run:

```
$ wget --post-data 'Australia, officially the Commonwealth of Australia, is a sovereign
country comprising the mainland of the Australian continent.'
'corenlp.run/?properties={"annotators":"pos","outputFormat":"json"}' -O -
```