Associated Press Ignite Talk 2 Contents
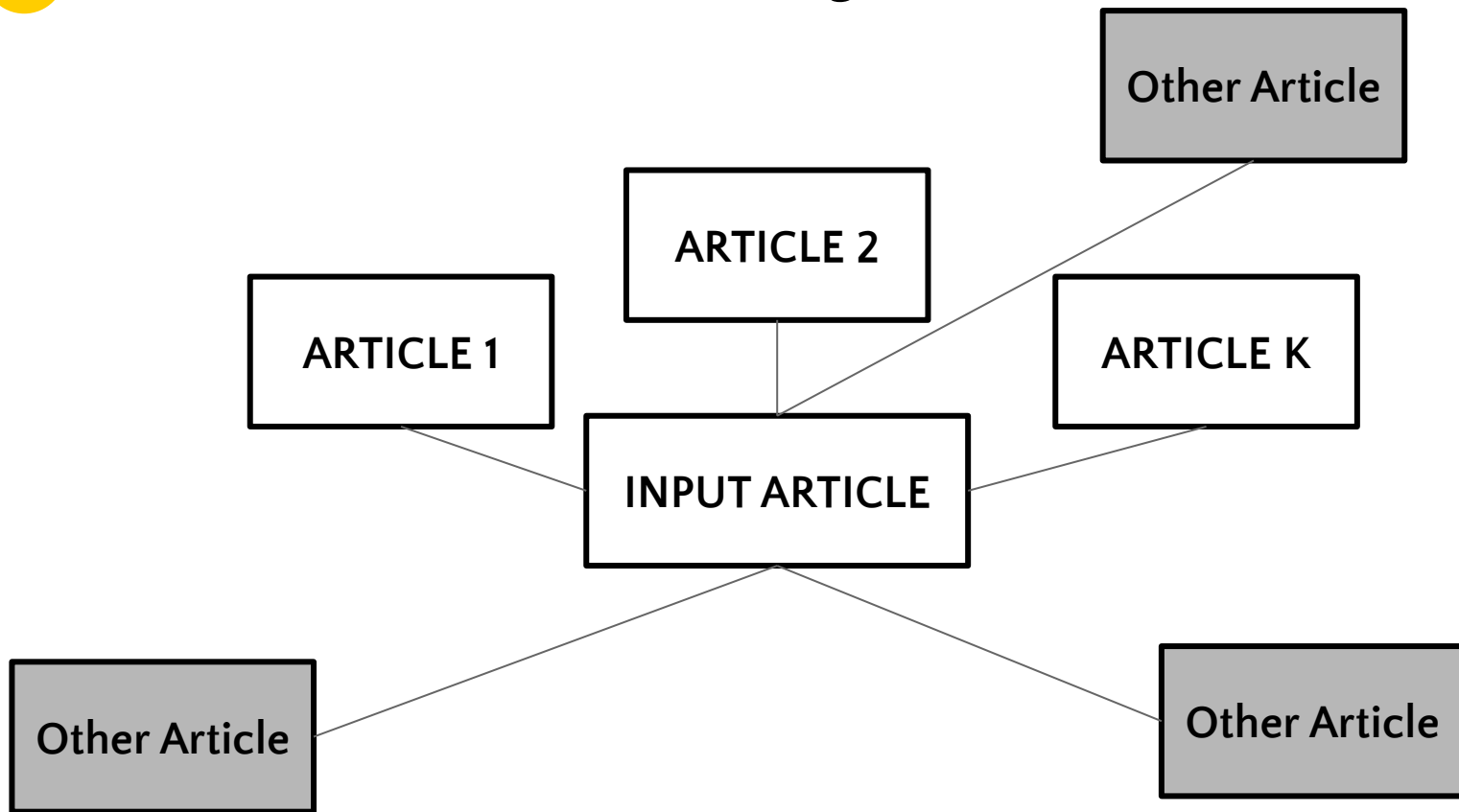
- Model
  - Baseline Model: kNN
  - Model Iteration I: Tag-to-Tag Recommendation
    - Article Tag Retrieval / Ranking:
      - Named Entities Recognition
      - TextRank
    - Image Tag Recommendation: Using Faster RCNN to get visual information
- UI and API Pipeline
  - UI predictions now made by hitting the API which calls a model

Baseline Model: K-Nearest Neighbors

Baseline Model: K-Nearest Neighbors

```
def KNN:
    # @param train: training set
    # @article_to_image: article to
        image map
    def fit(self, train,
article_to_image)

    # @param sim: fn to calculate
        similarity of tags
    # @param test_tags: test tags
set
    def predict(self, sim, test_tags)
```

```
test = ['sports', 'nba basketball', 'basketball' ,'professional
basketball', 'entertainment', 'nba playoffs'], k = 3
```



```
# normalized exact match
def baseline_score(train, test):
    # exact tag overlap / len(train)
```

Baseline Model: K-Nearest Neighbors

```
def KNN:
    # @param train: training set
    # @article_to_image: article to
        image map
    def fit(self, train,
article_to_image)

    # @param sim: fn to calculate
        similarity of tags
    # @param test_tags: test tags
set
    def predict(self, sim, test_tags)
```
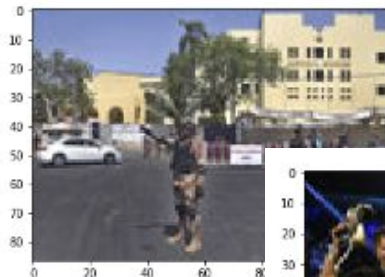
```
test = ['sports', 'nba basketball', 'basketball' ,'professional
basketball', 'entertainment', 'nba playoffs'], k = 3
```
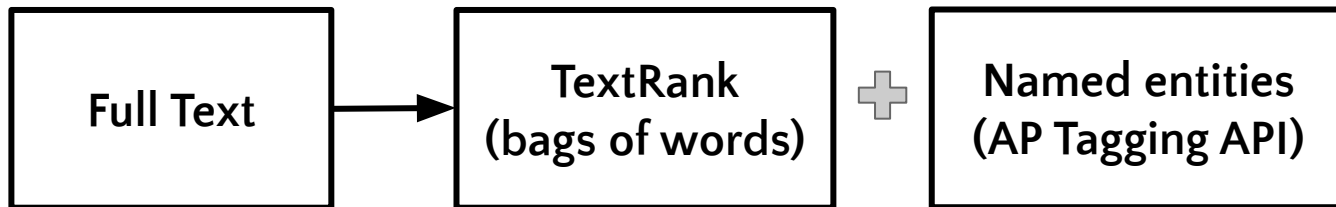


```
# normalized synonym match
def baseline_score(train, test):
    # exact tag overlap + eta * # similar tag overlap / len(train)
```

Article Tag Ranking: TextRank

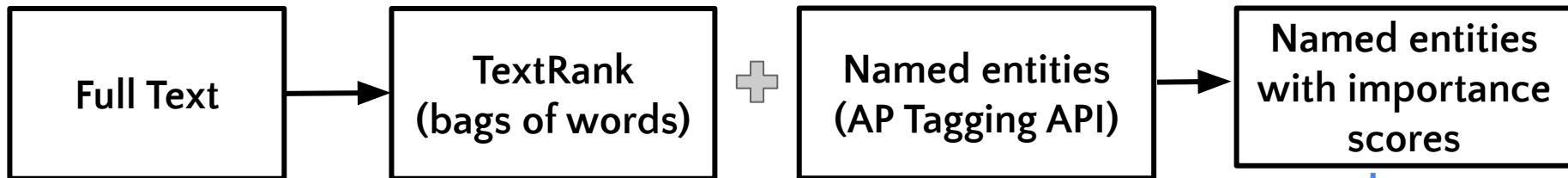| Full Text | → | TextRank (bags of words) | ✚ | Named entities (AP Tagging API) |

```
[('Party leader Kemal Kilicdaroglu', 0.047500269373415754),
 ('opposition party leader', 0.04491151046482916),
 ('Interior Minister Suleyman Soylu', 0.0393905668124247),
 ('President Recep Tayyip Erdogan', 0.03529180424934934),
 ('municipal election', 0.02023076940656441),
 ('kurdish rebel', 0.0201731337306259),
 ('mayoral election', 0.019878721600533583),
 ('Republican People', 0.019007513261570123),
 ('Turkey', 0.018045623982279352),
 ('divisive campaign', 0.014757984913261336)]
```

tag	type
General news	subject
Government and politics	subject
Funerals and memorial services	subject
Recep Tayyip Erdogan	person
Kemal Kilicdaroglu	person
Ankara	place
Turkey	place
Western Europe	place
Europe	place
Middle East	place
Turkey government	org

Article Tag Ranking: TextRank

Full Text → **TextRank (bags of words)** ➕ **Named entities (AP Tagging API)** → **Named entities with importance scores**

```
('Party leader Kemal Kilicdaroglu', 0.047500269373415754),
('opposition party leader', 0.04491151046482916),
('Interior Minister Suleyman Soylu', 0.0393905668124247),
('President Recep Tayyip Erdogan', 0.03529180424934934),
('municipal election', 0.02023076940656441),
('kurdish rebel', 0.0201731133730625 9),
('mayoral election', 0.019878721600533583),
('Republican People', 0.019007513261570123),
('Turkey', 0.018045623982279352),
('divisive campaign', 0.014757984913261336)]
```

tag	type
General news	subject
Government and politics	subject
Funerals and memorial services	subject
Recep Tayyip Erdogan	person
Kemal Kilicdaroglu	person
Ankara	place
Turkey	place
Western Europe	place
Europe	place
Middle East	place
Turkey government	org

TextRank Importance Score	Tagging API
0.0475003	Kemal Kilicdaroglu
0.0352918	Recep Tayyip Erdogan
0.0180456	Turkey

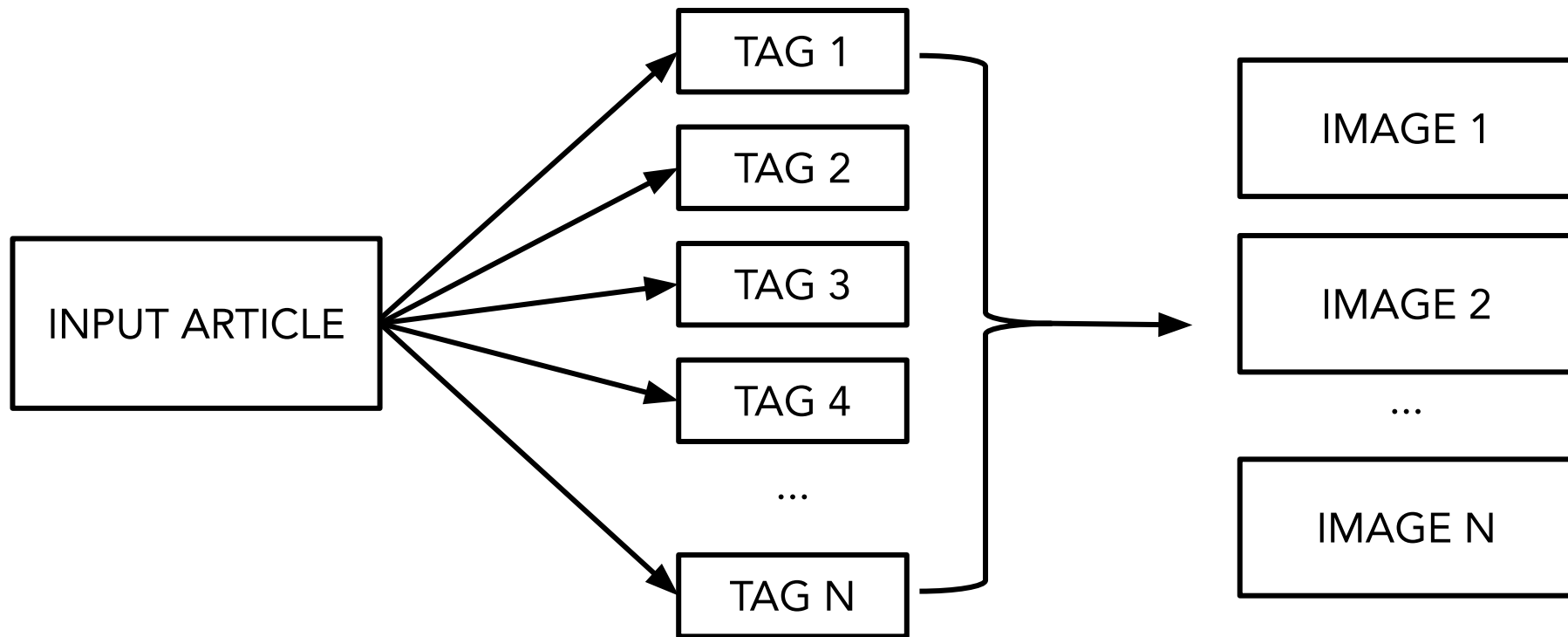# Model Iteration I: Tag-to-Tag Recommendation (Li et al., 2017)

Image Tag Importance Measurement (Li et al., 2017)

- Object Tag Importance (#7)

  ○ Discounted Probability Importance

- Scene Tag Importance (#8)

  ○ Whether a scene tag appears in description (#2)

    ■ Whether it appears as the modifier (#5)

    ■ or the subject (#6) of a description

Algorithm 1. Measuring Object/Scene Tag Importance in a Sentence Description

1: Input: The set of all object tags $\mathcal{T}_o$, the set of all object tags mentioned in sentence $\mathcal{T}_o^s$, current sentence s, scene tag t_s.
2: **Set** scene factor c_s to default value 0;
3: **Set** sentence tree **T** to parseSentence(s);
4: **Set** path **p** to findPath(**T**, **T**.$root$, t_s).
5: **If** "PP" in **p** and "PP".$left = prep$: **Set** $c_s = \alpha$;
6: **Else if** $\mathbf{p} \neq NULL$: **Set** $c_s = \beta$;
7: **For** $t \in \mathcal{T}_o$: $\mathbf{I}(t) = \frac{\mathbb{I}\left\{t \in \mathcal{T}_o^s\right\}}{|\mathcal{T}_o^s|(1+c_s)}$;
8: $\mathbf{I}(t_s) = \frac{c_s}{1+c_s}$.

Image Tag Importance Measurement

Caption:
['Pope Francis exchange gifts with South Sudan President Salva Kiir Mayardit during a private audience at the Vatican, Saturday, March 16, 2019. (AP Photo/Alessandra Tarantino, Pool)']

Object Tag Importance:
salva kiir mayardit: 0.042
pope francis: 0.042

Scene Tag Importance:
south sudan: 0.500
italy: 0.000
western europe: 0.000
europe: 0.000
north africa: 0.000
africa: 0.000
milan: 0.000
vatican city: 0.000

Caption:
['Pope Francis arrives for a audience with members of the confederation of the Italian cooperatives in the Paul VI Hall at the Vatican, Saturday, March 16, 2019. (AP Photo/Alessandra Tarantino,)']

Object Tag Importance:
pope francis: 0.250

Scene Tag Importance:
milan: 0.000
italy: 0.000
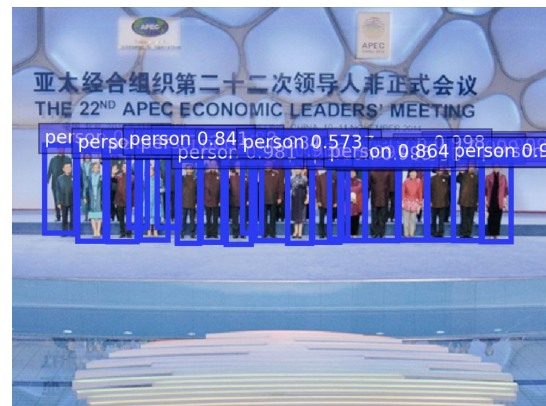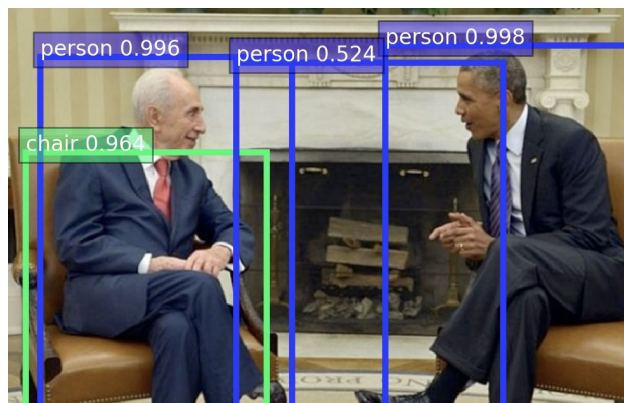western europe: 0.000
europe: 0.000
vatican city: 0.000
christianity: 0.000
religion: 0.000
social affairs: 0.000

Using Visual Cues

- Aim: Identify bounding boxes for the people in the image
- Importance of the person tag determined by the **size of the box**, **number of boxes** and **distance from the center** of the image
- This is done through a pre-trained Faster RCNN network

Article Input

Police Pursuit in Salt Lake City

The Latest on a police pursuit in Salt Lake City (all times local):5:20 p.m.Utah authorities say a man killed after leading police on a chase down a busy street is 37-year-old man from the Salt Lake City suburb of West Valle

Search

Input by the user

Hits an API endpoint that calls the tagging service and the model's 'predict' function

Tags

general news
police
law enforcement agencies

Similar Article Headlines

Officer of the Year relishes working busy bar break shift
US won't immediately give away $1B for California rail
The Latest: Utah man killed in police pursuit identified

Similar articles: shown when the model is kNN

Recommended Images

Recommended images shown. **Next steps:** show image metadata and ground truth images (if any)