

Learning to Read Academic Paper

Yining Hong, Jialu Wang, Xueheng Zhang, Zheng Wu

May 26th, 2018

Why

Why

What

Why What

How

What How
Motivation

How

Motivation

Service

Motivation Service Pipeline

Overview

Motivation

Service

Pipeline

1 st

or

Motivation

Why we need
“Learning to Read Academic Paper”?

arXiv

- Nowadays **tens of thousands** of papers released on arXiv every day
- Not all of them are of high quality



Fig 1. arXiv submission rate

2nd

UIC

Service

Service

Acemap

WeChat

Acemap

WeChat

3rd

CIC

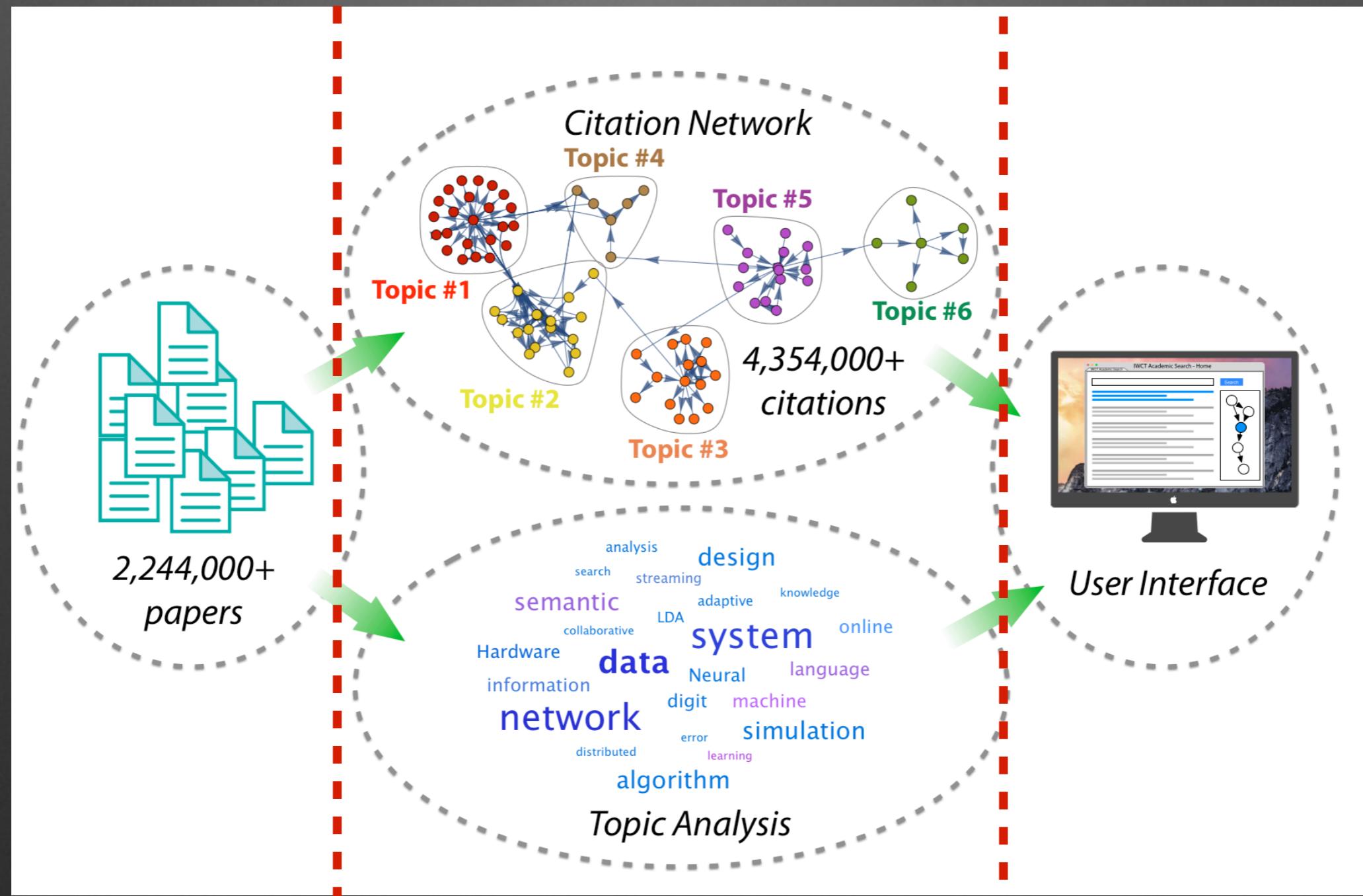
Pipeline

Pipeline

Data

Model

Visualization



Data

Acemap
Dataset

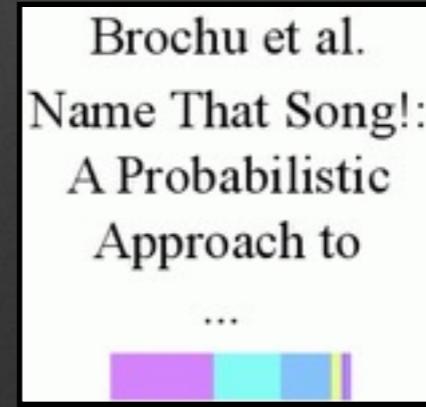
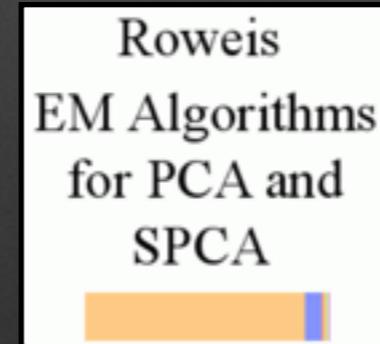
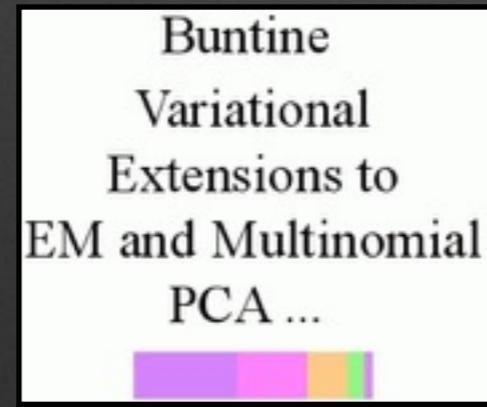
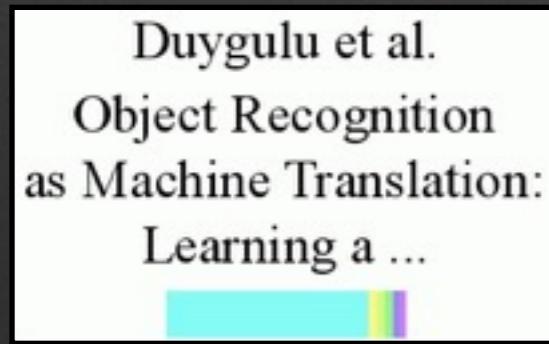
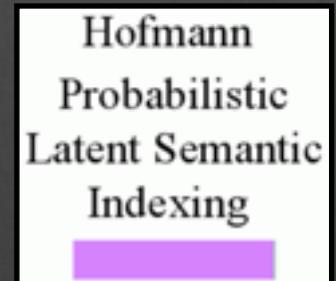
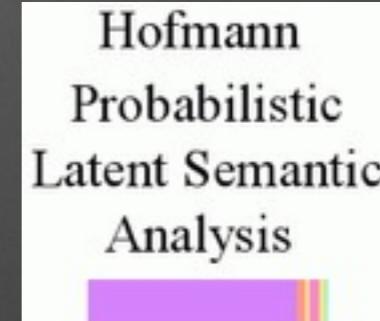
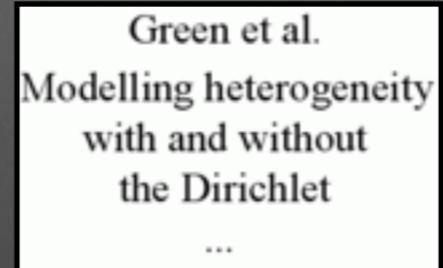
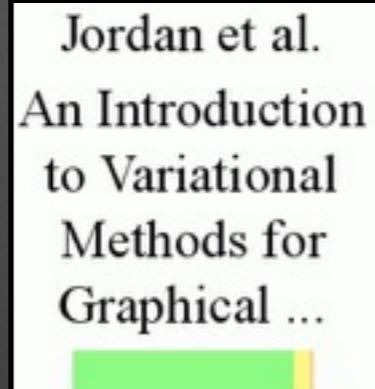
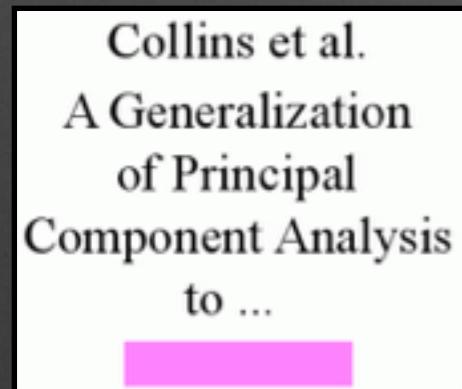


Python Crawler



Model

- FastText – View each paper individually



Model

Hofmann
Probabilistic
Latent Semantic
Indexing

Green et al.
Modelling heterogeneity
with and without
the Dirichlet
...

Jordan et al.
An Introduction
to Variational
Methods for
Graphical ...

Collins et al.
A Generalization
of Principal
Component Analysis
to ...

Roweis
EM Algorithms
for PCA and
SPCA

Blei et al.
Latent Dirichlet
Allocation

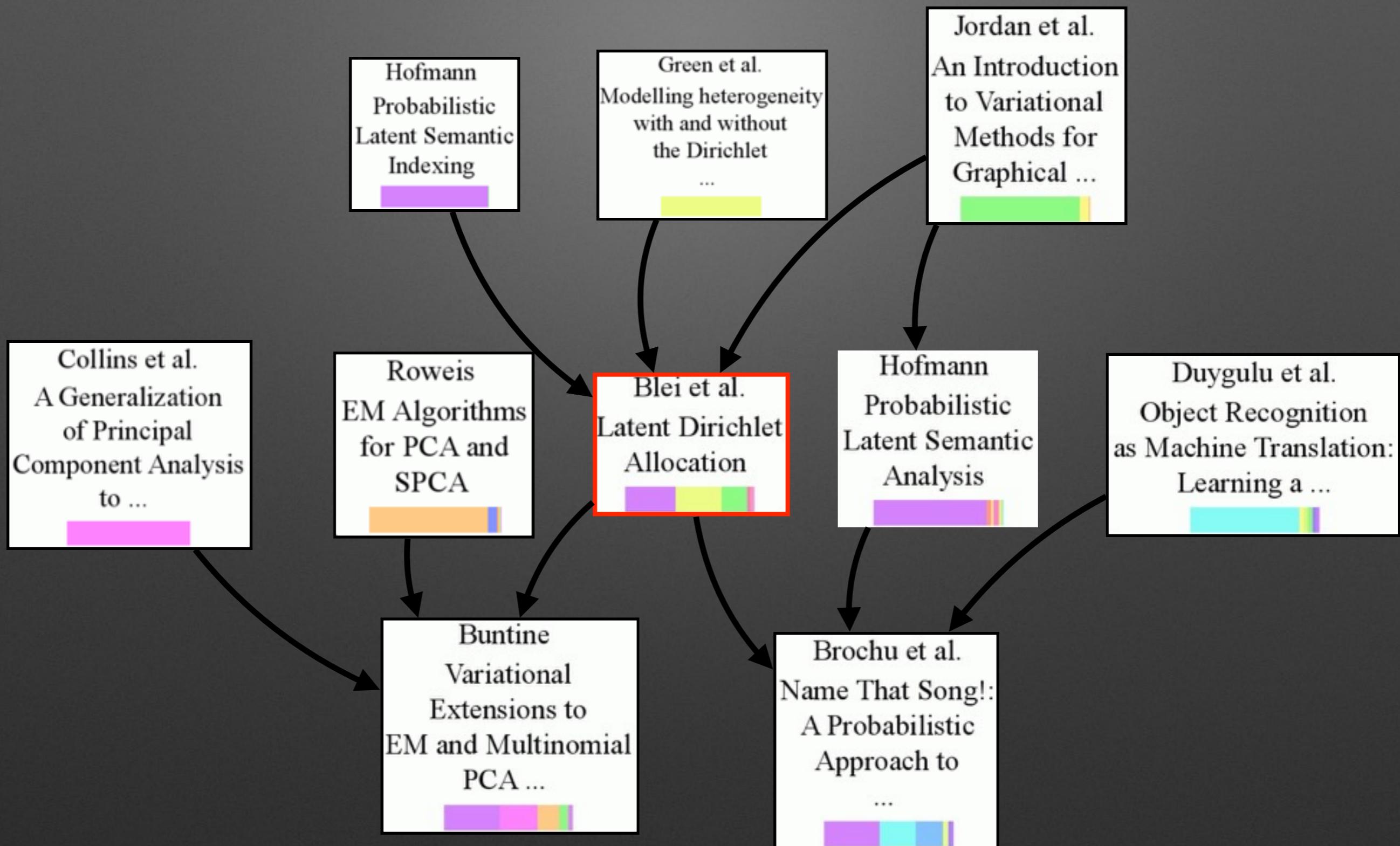
Hofmann
Probabilistic
Latent Semantic
Analysis

Duygulu et al.
Object Recognition
as Machine Translation:
Learning a ...

Buntine
Variational
Extensions to
EM and Multinomial
PCA ...

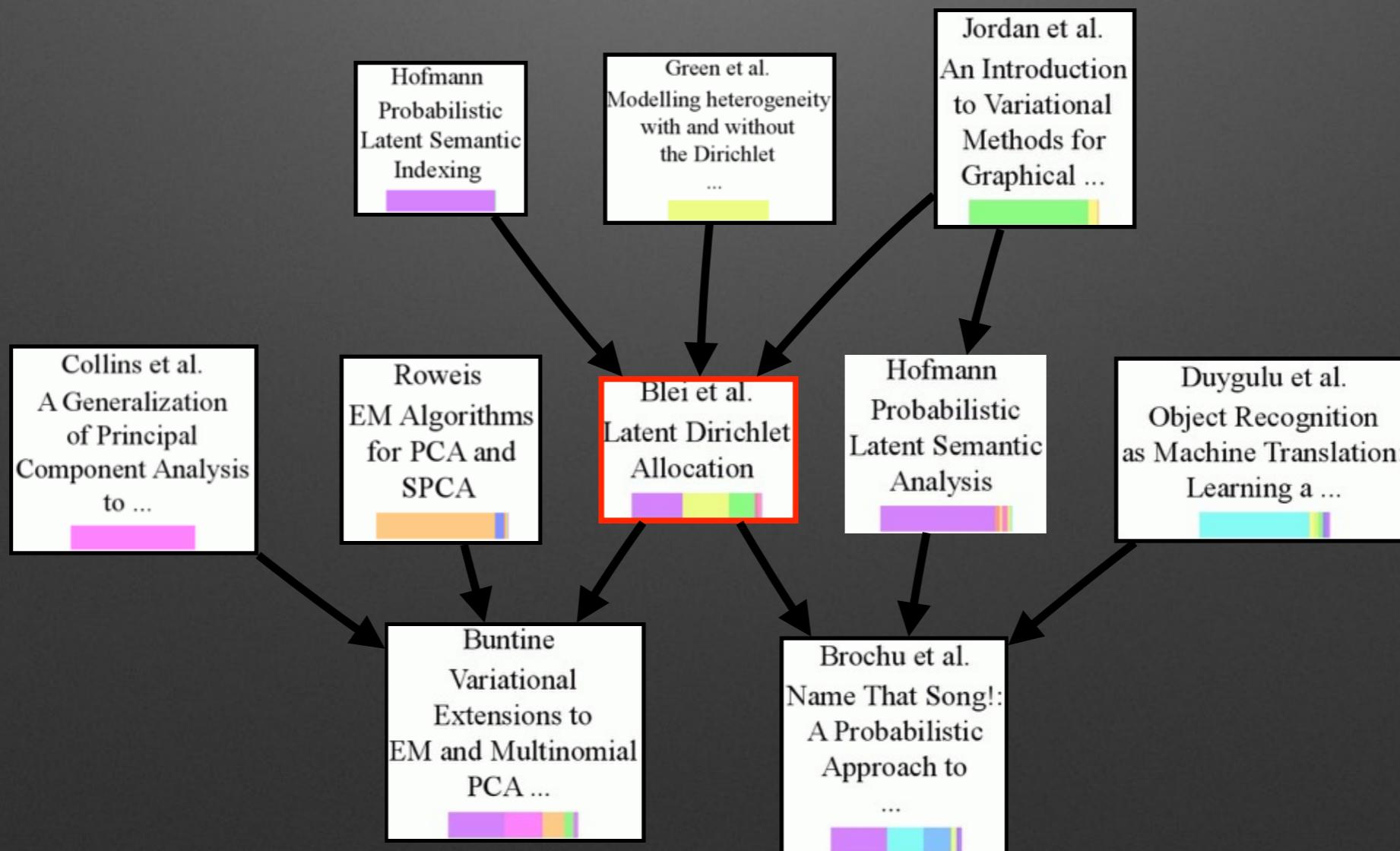
Brochu et al.
Name That Song!:
A Probabilistic
Approach to
...

Model



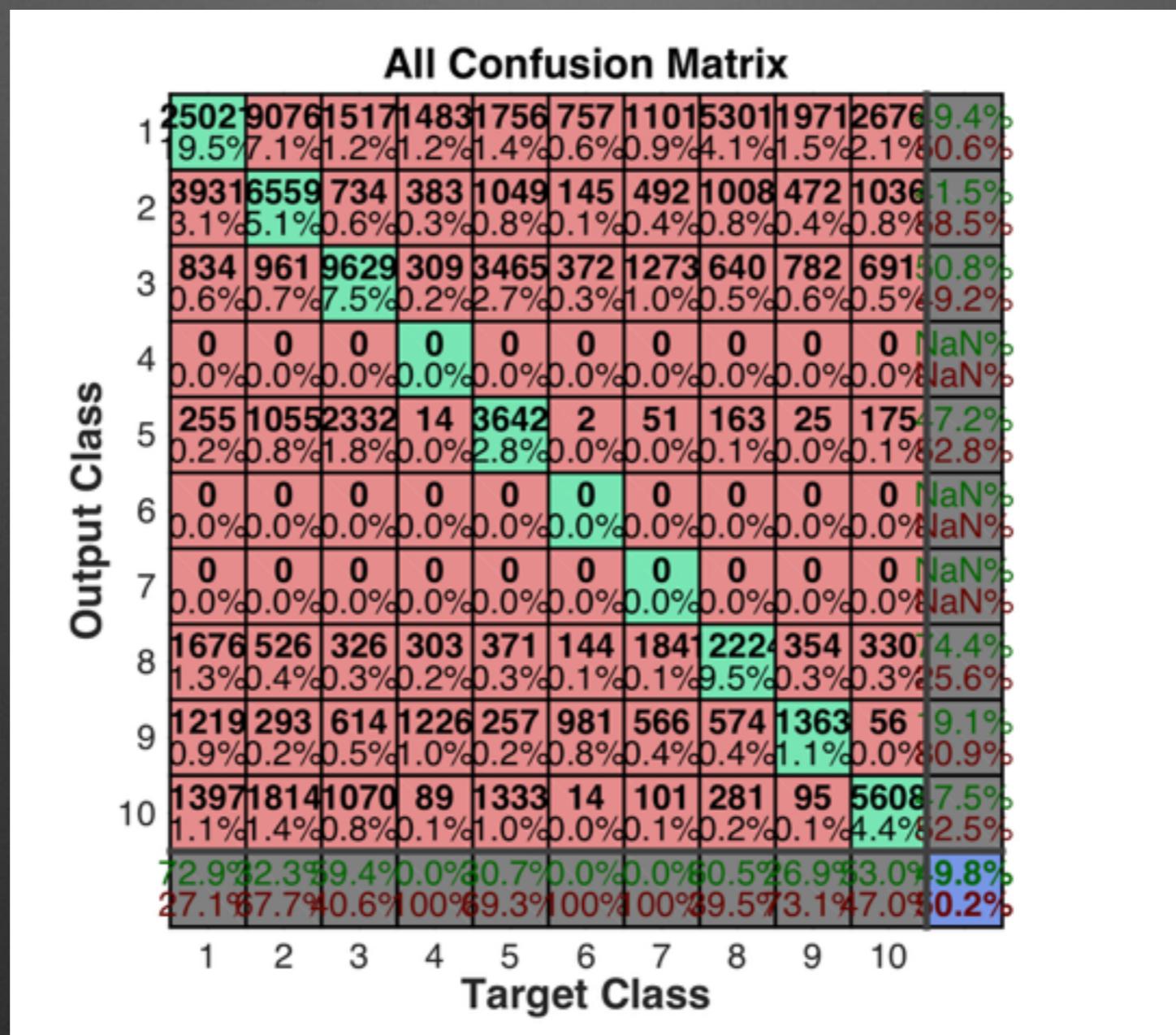
Model

- Citation Network— Analyze relationship among papers



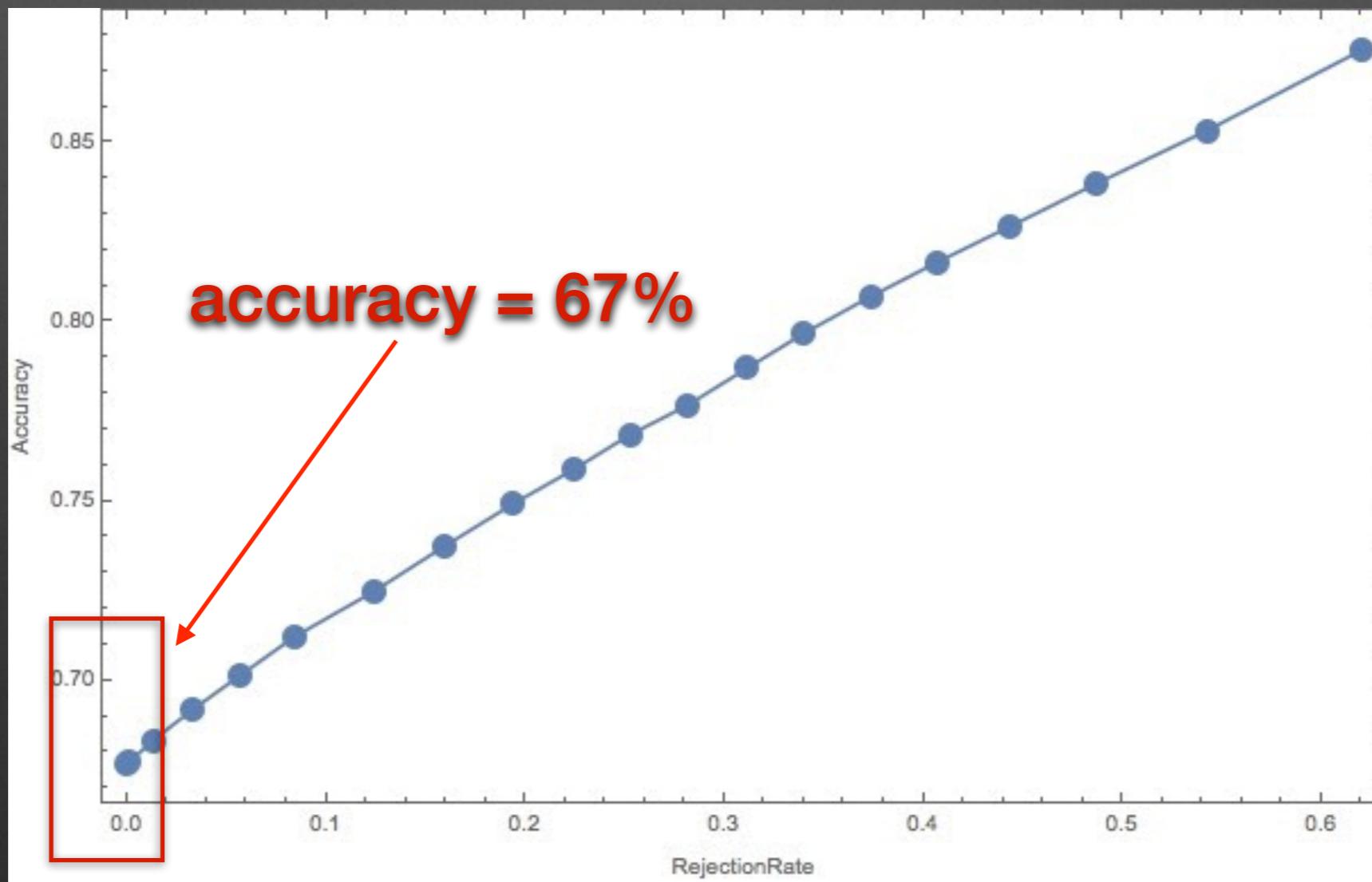
Model

- LDA + neural network



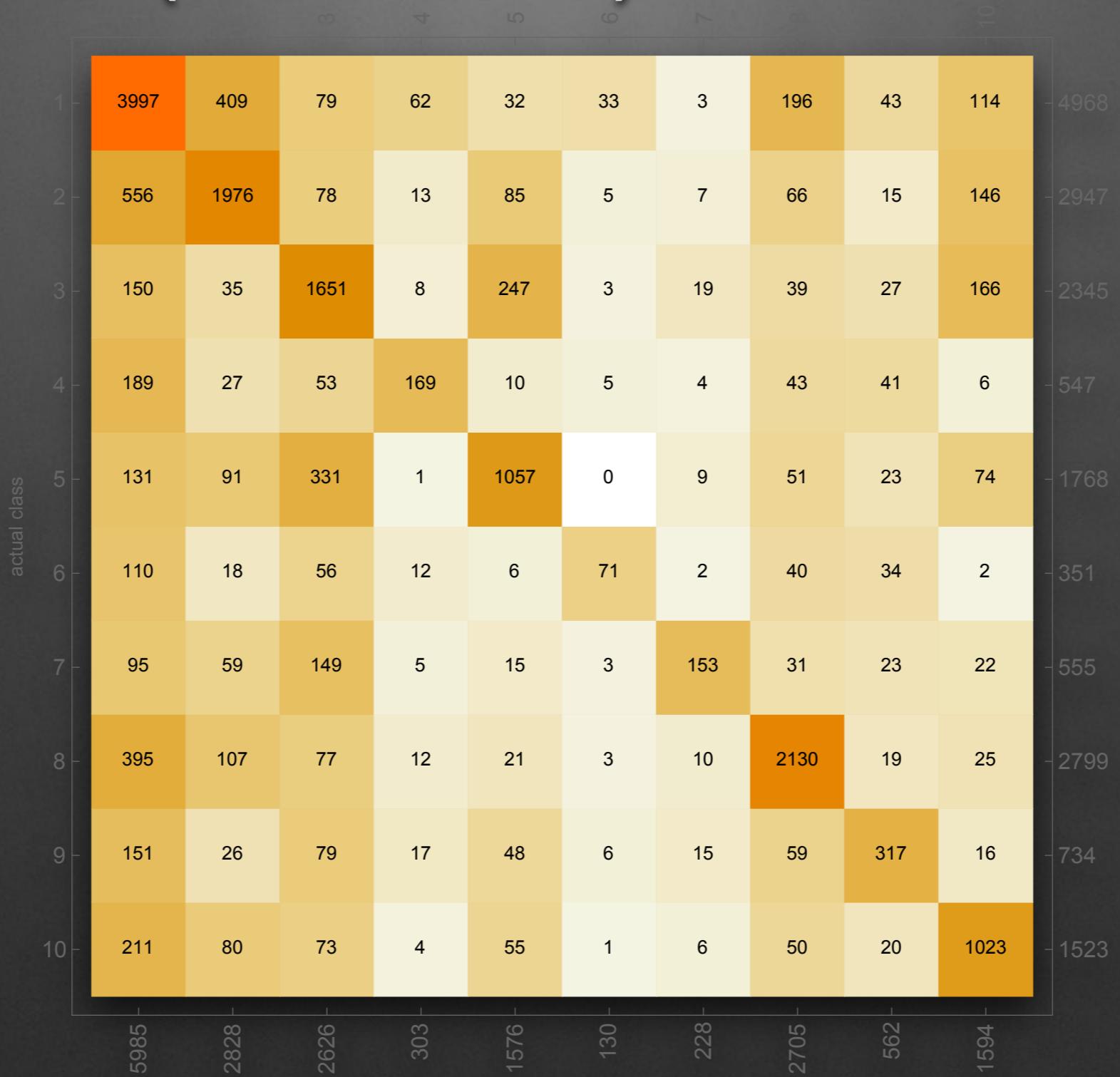
Model

- Markov Chain (on title words)
 - Mathematica built-in classifier



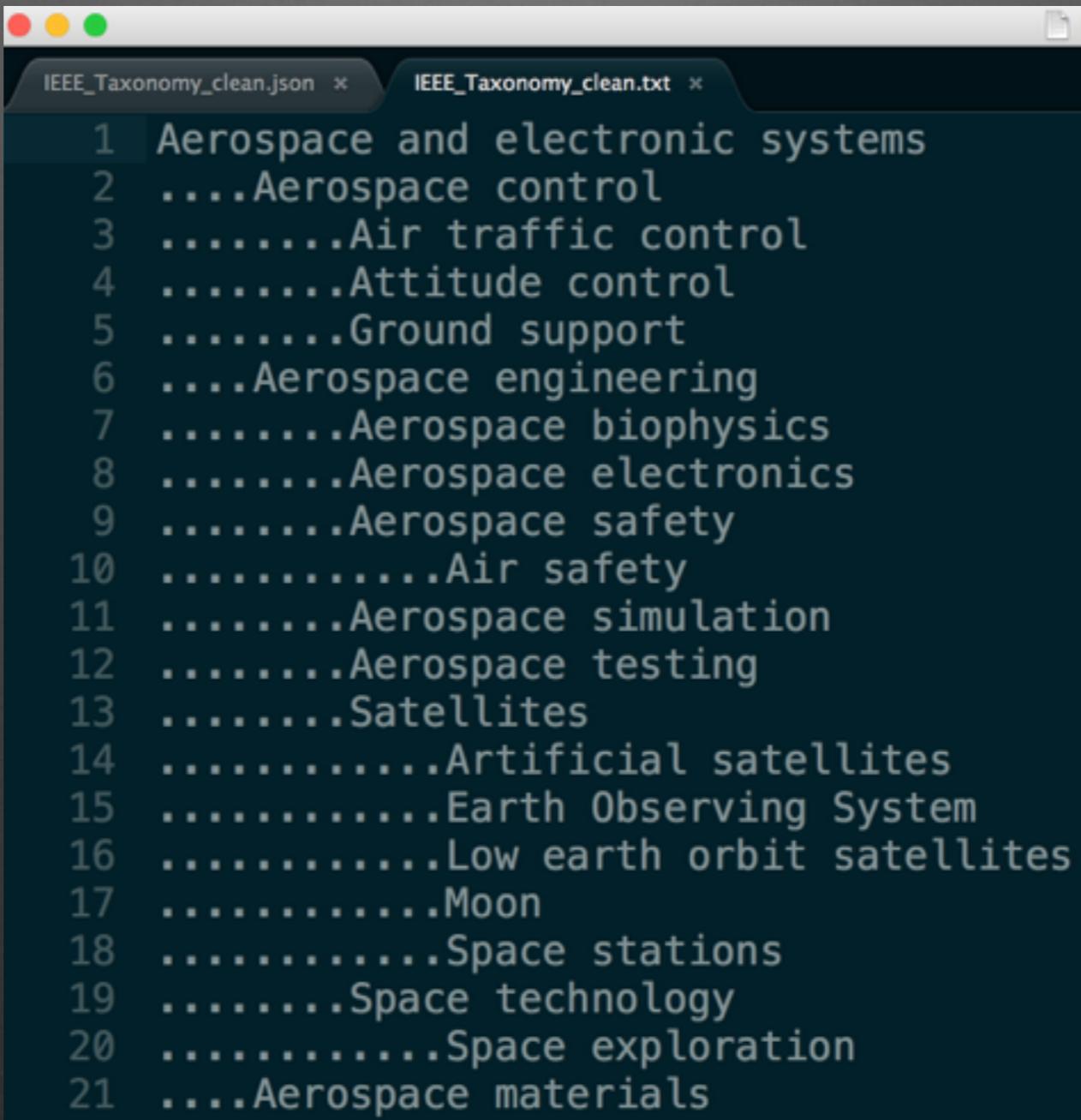
Model

- Markov Chain (on title words)



Model

- IEEE taxonomy (TXT Format)

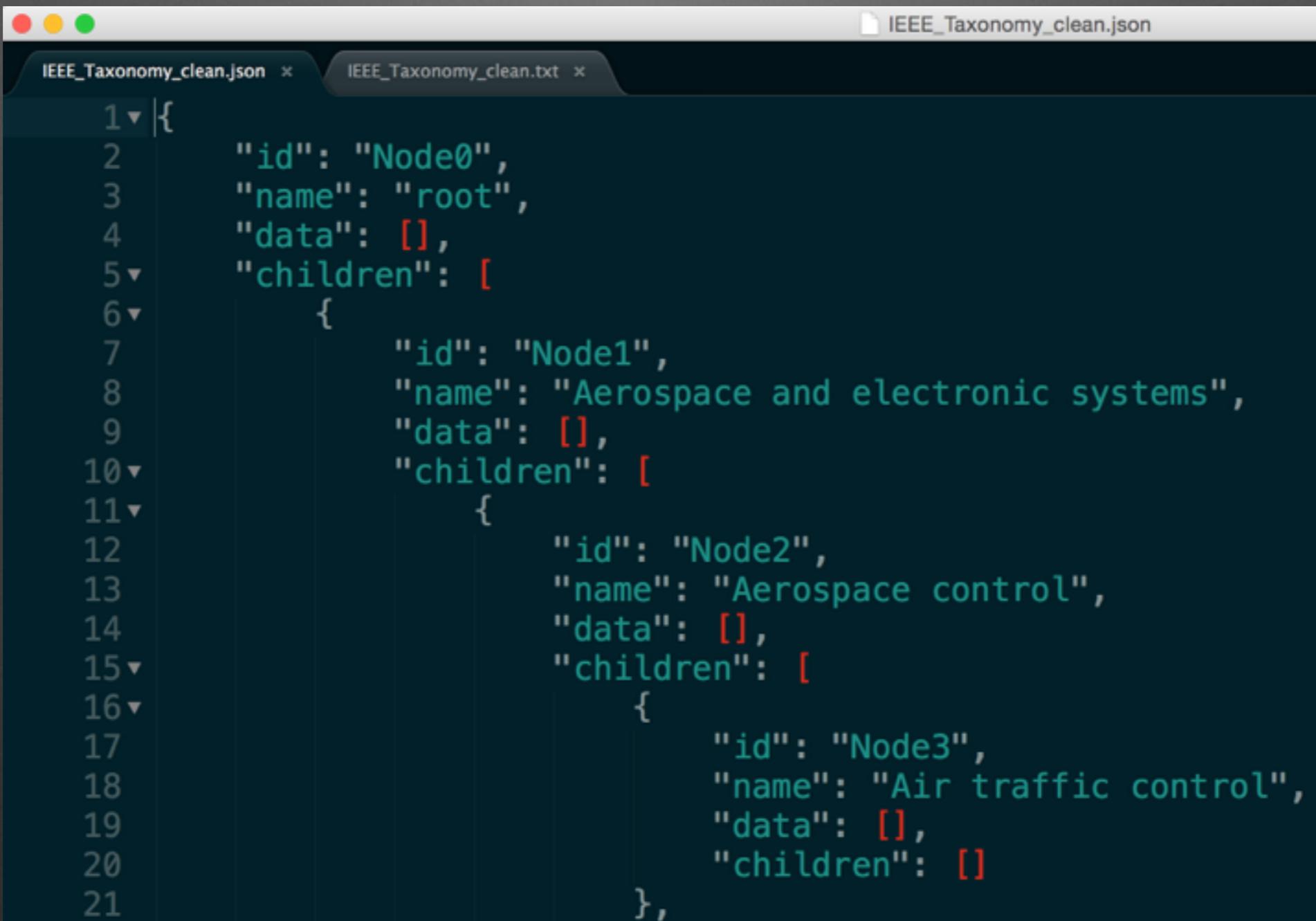


The image shows a terminal window with two tabs: "IEEE_Taxonomy_clean.json" and "IEEE_Taxonomy_clean.txt". The "IEEE_Taxonomy_clean.txt" tab is active, displaying a hierarchical list of aerospace categories. The categories are numbered from 1 to 21, with each number followed by a category name and a varying number of dots indicating its level of nesting.

Number	Category
1	Aerospace and electronic systems
2Aerospace control
3Air traffic control
4Attitude control
5Ground support
6Aerospace engineering
7Aerospace biophysics
8Aerospace electronics
9Aerospace safety
10Air safety
11Aerospace simulation
12Aerospace testing
13Satellites
14Artificial satellites
15Earth Observing System
16Low earth orbit satellites
17Moon
18Space stations
19Space technology
20Space exploration
21Aerospace materials

Model

- IEEE taxonomy (XML Format)

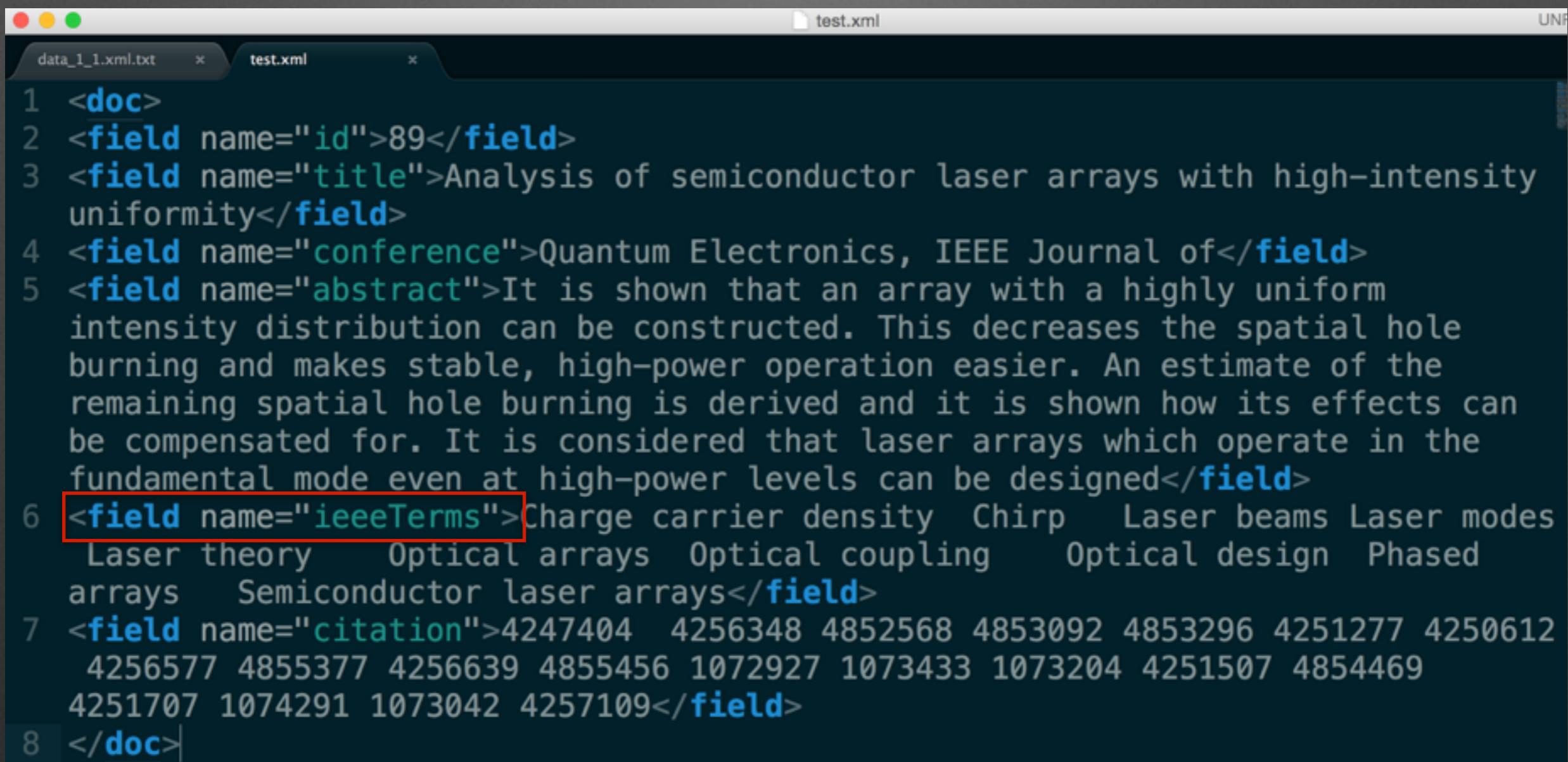


The screenshot shows a terminal window with two tabs: "IEEE_Taxonomy_clean.json" and "IEEE_Taxonomy_clean.txt". The "IEEE_Taxonomy_clean.json" tab is active, displaying a hierarchical JSON object representing an IEEE taxonomy. The JSON structure starts with a root node "Node0" and its children, which are further subdivided into "Node1", "Node2", and "Node3". The "name" field for each node contains descriptive text, such as "root", "Aerospace and electronic systems", "Aerospace control", and "Air traffic control". The "children" field for each node is an empty array, indicated by the red color of the brackets.

```
1 {  
2     "id": "Node0",  
3     "name": "root",  
4     "data": [],  
5     "children": [  
6         {  
7             "id": "Node1",  
8             "name": "Aerospace and electronic systems",  
9             "data": [],  
10            "children": [  
11                {  
12                    "id": "Node2",  
13                    "name": "Aerospace control",  
14                    "data": [],  
15                    "children": [  
16                        {  
17                            "id": "Node3",  
18                            "name": "Air traffic control",  
19                            "data": [],  
20                            "children": []  
21                        },  
22                    ],  
23                ],  
24            ],  
25        ],  
26    ],  
27}
```

Model

- IEEE keywords



```
1 <doc>
2 <field name="id">89</field>
3 <field name="title">Analysis of semiconductor laser arrays with high-intensity uniformity</field>
4 <field name="conference">Quantum Electronics, IEEE Journal of</field>
5 <field name="abstract">It is shown that an array with a highly uniform intensity distribution can be constructed. This decreases the spatial hole burning and makes stable, high-power operation easier. An estimate of the remaining spatial hole burning is derived and it is shown how its effects can be compensated for. It is considered that laser arrays which operate in the fundamental mode even at high-power levels can be designed</field>
6 <field name="ieeeTerms">Charge carrier density Chirp Laser beams Laser modes  
Laser theory Optical arrays Optical coupling Optical design Phased  
arrays Semiconductor laser arrays</field>
7 <field name="citation">4247404 4256348 4852568 4853092 4853296 4251277 4250612  
4256577 4855377 4256639 4855456 1072927 1073433 1073204 4251507 4854469  
4251707 1074291 1073042 4257109</field>
8 </doc>
```

Visualization

Visualization

- Solr – Open source search engine
- Amaze UI – HTML5 Web page framework
- JIT – JavaScript InfoVis ToolKit



Visualization

- Deploy Apache on our server
- Install and configure Solr
- Develop web front-end pages (Amaze UI)
- Transform all dataset to required formats (json, xml)
- Run Solr and return search results to specific webpage
- Add animation with JIT

Visualization

AceRec

Get Started

Topic List Result

Help Login

ACADEMIC PAPER RECOMMENDATION

请输入你想找的话题

Search

More info

By Author:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

By Topic:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

By Published Year:

Before 1950 1950s 1960s 1970s 1980s 1990s 2000–2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

Visualization

By Topic:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

By Published Year:

Before 1950 1950s 1960s 1970s 1980s 1990s 2000–2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

Statistics

of publications: 2,909,788
of authors: 1,537,934
of conferences: 3,759
of journals: 1,409

Topic Display

FEATURE 1

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

FEATURE 2

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

FEATURE 3

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

Visualization

[Network servers](#)

[Network servers](#)

[Network topology](#)

[Network interfaces](#)

[Network synthesis](#)

[Network servers](#)

[PACE network](#)

[Nonlinear network analysis](#)

[Nonlinear network analysis](#)

[Computer network management](#)

Orientation:

Max levels:

[Go to Parent](#)

[See the Example Code](#)

Communications technology

Communication equipment

Communication switching

Communication systems

Couplers

High-speed electronics

Image communication

Message systems

Modulation

Multiplexing

Network topology

Presence network agents

TV

UHF technology

Ultra wideband technology

VHF devices

Directional couplers

Review

- Motivation
- Service
 - Acemap, WeChat
- Pipeline
 - Data / Model / Visualization

Thank you

Q & A