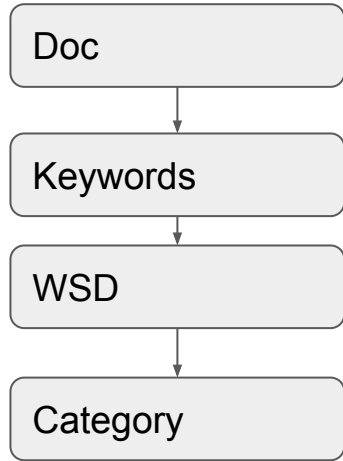# Youtube Documentary Videos

- Mandar Kakade
- Manoj Preveen

Basic Approach:
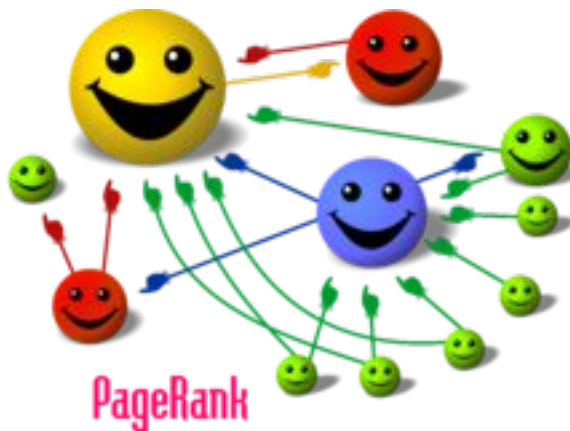
COMPLETE PATH :

Doc → Keywords → WSD → Category

Subtitle file → Tokenize → Words → Stop words Removed → **Tokens** → POS Tagging → **Annotated Text**

Annotated Text → Keyword Extraction → **Keywords**

Keywords → Word Sense Disambiguation → **Keywords in correct sense** → WordNet → **Synsets in correct sense**

Synsets in correct sense → **Domains** → Domain with Highest Occurrence Score → **Category**
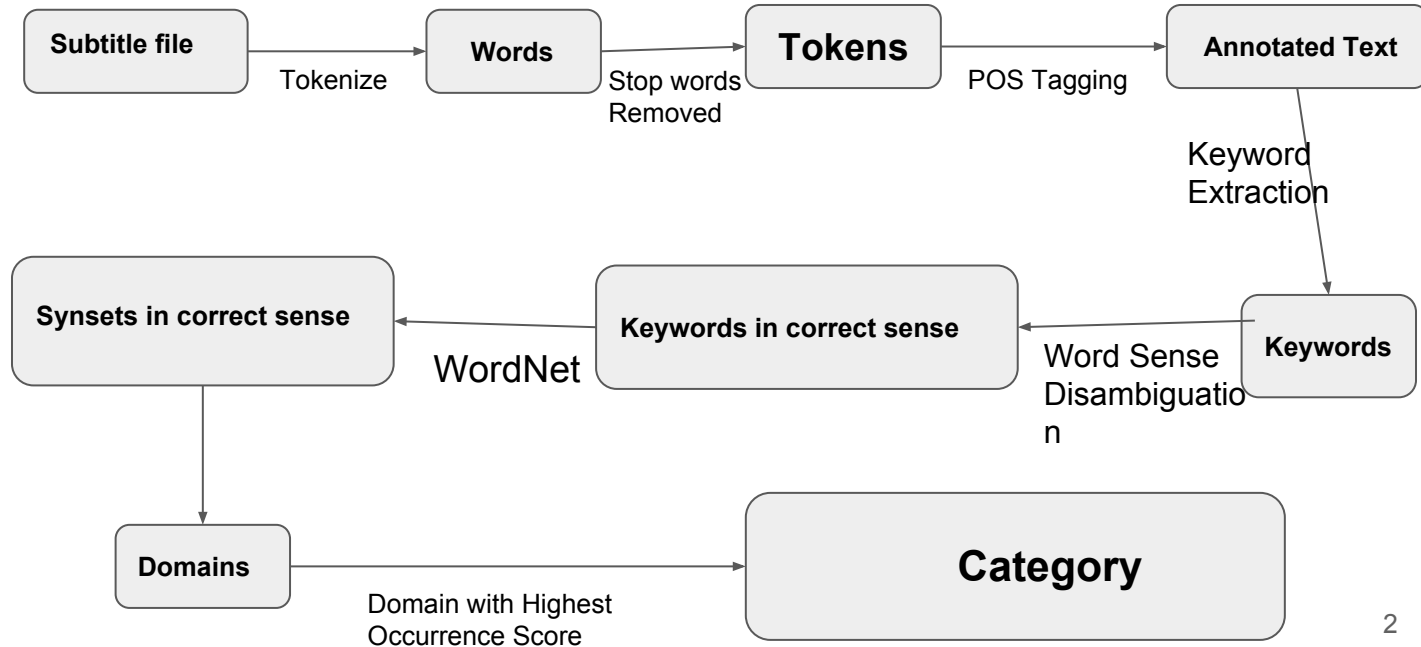
PageRank Algorithm

TextRank Algorithm

If docs T1….Tn point to A.

$$PR(A) = (1-d) + d(PR(T1)/C(T1) + \cdots + PR(Tn)/C(Tn))$$

Compatibility of systems of linear constraints over the set of natural numbers. Criteria of compatibility of a system of linear Diophantine equations, strict inequations, and nonstrict inequations are considered. Upper bounds for components of a minimal set of solutions and algorithms of construction of minimal generating sets of solutions for all types of systems are given. These criteria and the corresponding algorithms for constructing a minimal supporting set of solutions can be used in solving all the considered types systems and systems of mixed types.
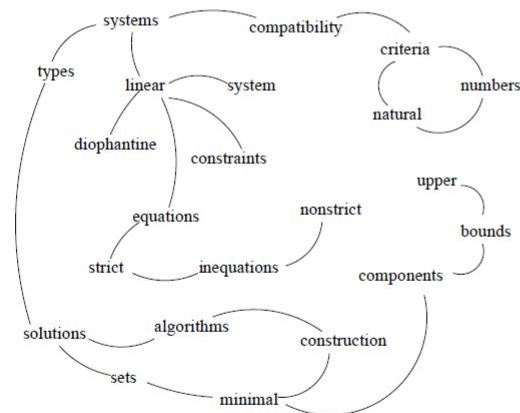
Keywords assigned by TextRank:
linear constraints; linear diophantine equations; natural numbers; nonstrict inequations; strict inequations; upper bounds

Keywords assigned by human annotators:
linear constraints; linear diophantine equations; minimal generating sets; non-strict inequations; set of natural numbers; strict inequations; upper bounds

Lesk's Algorithm

```
>>> for i in wn.synsets('bank'):
        print(i.definition())

sloping land (especially the slope beside a body of water)
a financial institution that accepts deposits and channels the money into lending activities
```

BANK

Financial Bank?

River Bank?

The bank of river Ganga is beautiful.

I borrowed money from a bank.

Domain of a word changes with sense.

EXAMPLE: BEFORE THE FLOOD

1. INPUT: Subtitle files in '.vtt' format

```
WEBVTT
Kind: captions
Language: en
Style:
::cue(c.colorCCCCCC) { color: rgb(204,204,204);
 }
::cue(c.colorE5E5E5) { color: rgb(229,229,229);
 }
##

00:00:03.020 --> 00:00:08.450 align:start position:0%

we've<00:00:07.790><c> known</c><00:00:08.059><c> about</c>
<00:00:08.120><c> this</c><00:00:08.450><c> for</c>

00:00:08.450 --> 00:00:08.570 align:start position:0%
we've known about this for

00:00:08.570 --> 00:00:11.030 align:start position:0%
we've known about this for
decades<c.colorE5E5E5><00:00:08.990><c> for</c></c><00:00:10.340>
<c> over</c><00:00:10.610><c> half</c><00:00:10.700><c> a</c>
<00:00:11.030><c> century</c>

00:00:11.030 --> 00:00:11.059 align:start position:0%
```

EXAMPLE: BEFORE THE FLOOD

2. Corresponding text file:

we've known about this for decades for over half a century try to have a conversation with anyone about climate change people just to now climate change climate change and the problem seems to be getting worse and worse worth the truth is the more I've learned about this issue and everything that contributes to the problem the more I realize how much I don't know oh my God look out violent that is paradise that has been degraded and destroyed or knowingly doing I just want to know how far we've gone and if there's anything we can do to stop it the US has been the biggest emitter of greenhouse gases in history your fossil addicted country we are doing more investment in solar today than the US half the people in office still don't believe in climate change why do you think there is such opposition about the science i think it's politics people are so arrogant to think they can change climate environmental issues have become the biggest reason for mass demonstrations we want to hold them accountable this is the most important issue of our time question is can we change our course in time you need a hundred get back to transition into energy that would make the United States for what the whole world all energy a sense that people all that I've seen on my journey shows us we have the means of stopping this devastation politicians do what the people want them to do what the American people are convinced the politicians will fall in line very quickly if we keep pushing there's no reason why we can't solve this problem the world is now watching we ask you to protect me are we and all living things we cherish our history yeah

EXAMPLE: BEFORE THE FLOOD

3. Keywords:

4. Category:

ENVIRONMENT.

think opposition
God
history yeah
anyone climate
energy sense
half century
time question
world energy
anything
destroyed
arrogant think
course time
accountable
science think
emitter greenhouse
opposition science
investment solar
solar today
today US
much
change problem
change think
devastation
reason mass

devastation
reason mass
important issue
conversation anyone
truth
journey
issue everything
United
country
issue time
whole world
transition energy
change course
US half
American
office
line
climate change
history fossil
change climate
US
ask
climate environmental
violent paradise

Results:

- Correctness of the algorithm is about 30%.

- The inputs for the algorithm were auto-generated subtitles and hence contain a lot of errors.
- Also videos were very short (2 min).
- Thus improving the input can greatly increase correctness.

Thank You!