

*IASC Diee*



# ***Studying the Impact of Text Summarization on Contextual Advertising***

G. Armano, **A. Giuliani**, and E. Vargiu

Intelligent Agents and Soft-Computing Group

Dept. of Electrical and Electronic Engineering

University of Cagliari, Italy

email: [alessandro.giuliani@diee.unica.it](mailto:alessandro.giuliani@diee.unica.it)

# Outline

- Online Advertising and Contextual Advertising
- Text Summarization Methods
- Evaluation
- The Adopted Contextual Advertising System
- Experimental Results
- Conclusions



# Online Advertising

- Form of promotion that uses the WWW for delivering marketing messages to attract customers

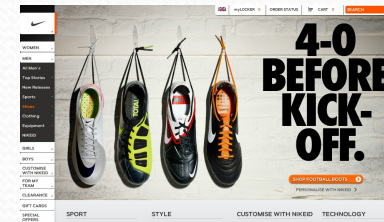


# Contextual Advertising

Web Page



Ad





# Contextual Advertising

The screenshot shows a fishing website with several AdSense advertisements. Red circles and arrows highlight the ads and their placement relative to the site's content. The ads include:

- Online Carp Fishing Bait**: Massive selection of carp fishing bait. Pellets, boilies, liquids. [www.bankside tackle.co.uk](http://www.bankside tackle.co.uk)
- Fishing**: Fish in those hard to get to places No trailers - go fishing anywhere [www.porta-bote.com](http://www.porta-bote.com)
- Simply Fishing**: Interested in Christian Videos? Watch Our Free Videos Online Today! [www.Inspiration.net](http://www.Inspiration.net)
- Ontario Flyin Fish Lodge**: Only lodge on 3 lakes and a river. American Plan. Very nice Lodge! [www.labelllake.com/](http://www.labelllake.com/)
- Carp Angling in Portugal**: Company offering fishing holidays and Carp baits for Carp Angling. [www.fishingportugal.com](http://www.fishingportugal.com)
- Quality Fishing Equipment**: Lures, Tuna, Tackles all in 1 place Tested and Guaranteed! [www.anglermania.com](http://www.anglermania.com)
- Snapper Fishing Experts**: melbourne snapper fishing Snapper Fishing port philip bay [www.melbournefishing.com.au](http://www.melbournefishing.com.au)
- Fly Fishing Conditions**: Up-to-Date Fly Fishing Conditions Reported by Local Outfitters [www.flyfishingconditions.com/](http://www.flyfishingconditions.com/)

The main content of the website includes a navigation menu, a search bar, and a section titled "General Tips" which discusses the history and popularity of fishing. A red arrow points from the "AdSense ads" label to the "Practice Good Stewardship of our Waterways:" section.

**AdSense ads**

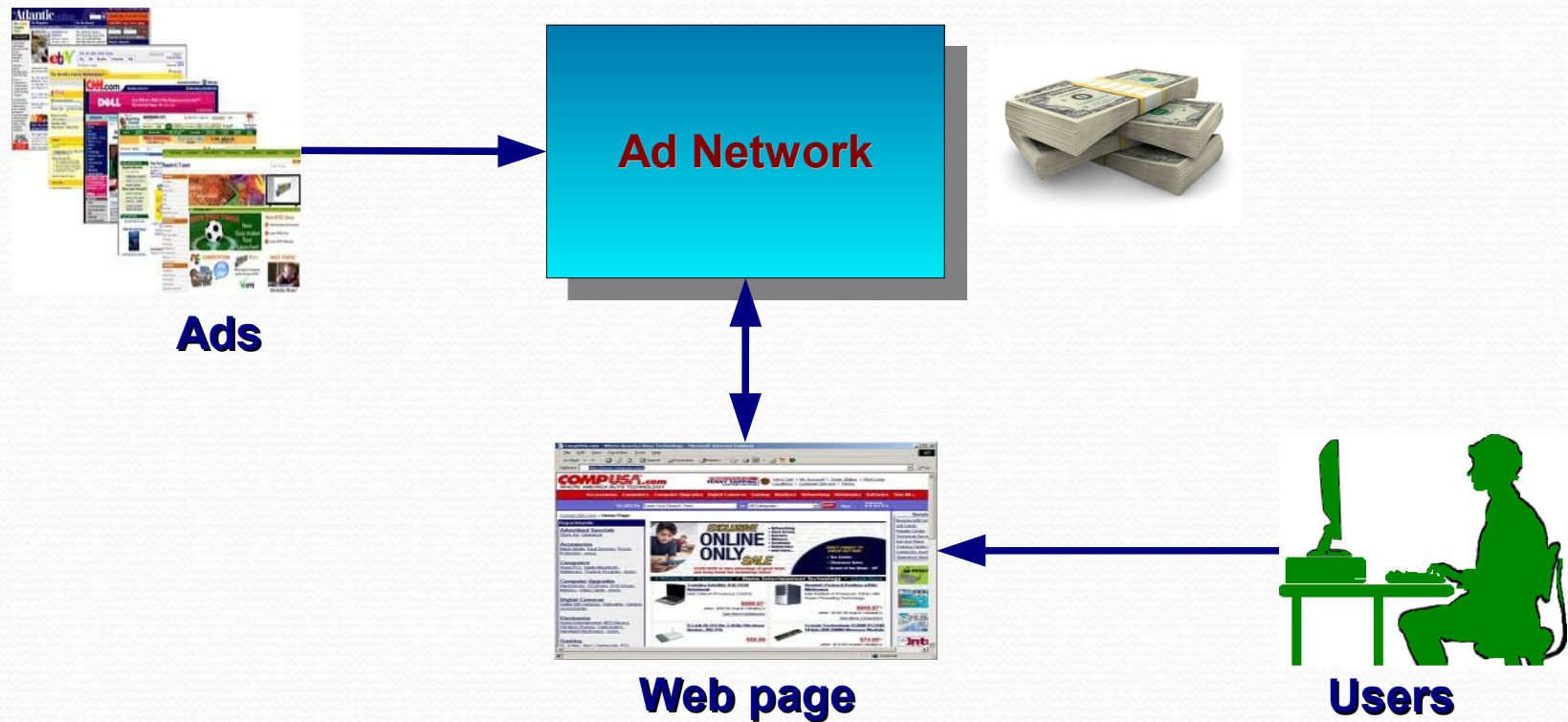
**Practice Good Stewardship of our Waterways:**

You can help to take care of our lakes, rivers, and other waterways so that others may enjoy these areas for years to come by practicing some of the following actions:

- Don't Litter... take along a trash bag or other receptacle for collecting your trash so that you can deposit it in the proper trash receptacle. Use proper dumping stations instead of tossing refuse into the water.
- Make sure that you use the correct type of bait and fishing gear permitted in that area. There may also be limits on the number, size, and kind of fish that you can keep. Check with your destination ahead of time to see what the local regulations allow. If you use a boat or watercraft when fishing, check to see what kinds of watercraft are allowed at the body of water where you are going to fish.
- Pay attention to local procedures and cautions for cleaning your watercraft after you leave the water so that you don't introduce the spread of nonnative species, such as the Zebra Mussel, to the next body of

# Contextual Advertising

- Task of placing ads within the content of a Web page





# Classic approaches

- High computation in data, resources, and time



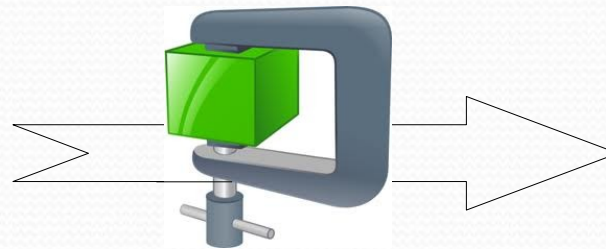
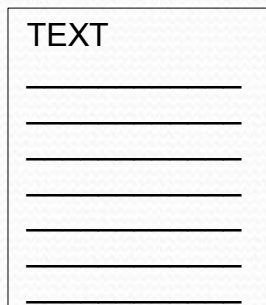
“Offline” advertising

# Current approaches



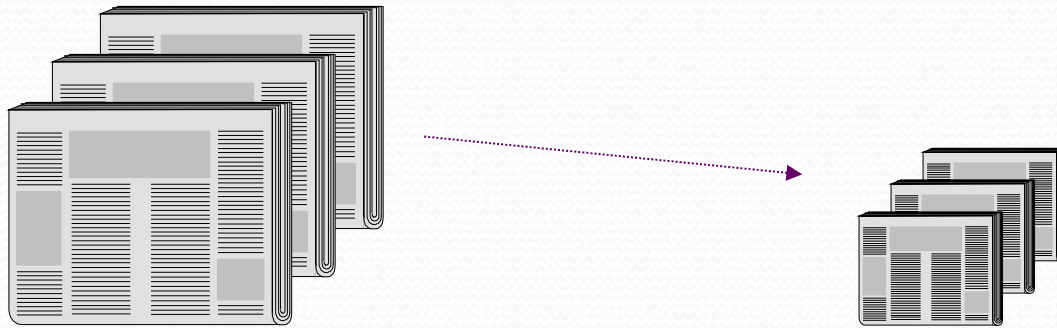
→ highly dynamic! → Real time advertising!

- Need of reducing data → text summarization





# *Focus: Text Summarization*



- Summaries:
  - obtained by single or multiple documents
  - preserve important information
  - short

# Focus: TS in CA

- **Extract vs. abstract**
  - lists fragments of text vs. re-phrases content coherently

in CA



- Extractive techniques!!!
  - Low computation
  - Reliance on single documents
  - Simple summaries (but effective)
  - Easier!



# *TS: Classic Approaches*

- Kolcz's methods → selection of meaningful paragraphs:
  - Title
  - First Paragraph
  - First and Second Paragraphs
  - First and Last Paragraphs
  - Paragraph with most Keywords
  - Paragraph with most Title Words

# TS in CA: Our Proposal



```
<!DOCTYPE  
<HTML>  
<HEAD>  
<TITLE>RA  
<LINK REV  
<META NAM
```

- Input: HTML code → need of additional features
  - Adoption of title of the web page!
- Our techniques:
  - Title and First Paragraph
  - Title and First Two Paragraphs
  - Title, First and Last Paragraphs
  - Most Title Words and Keywords Paragraphs

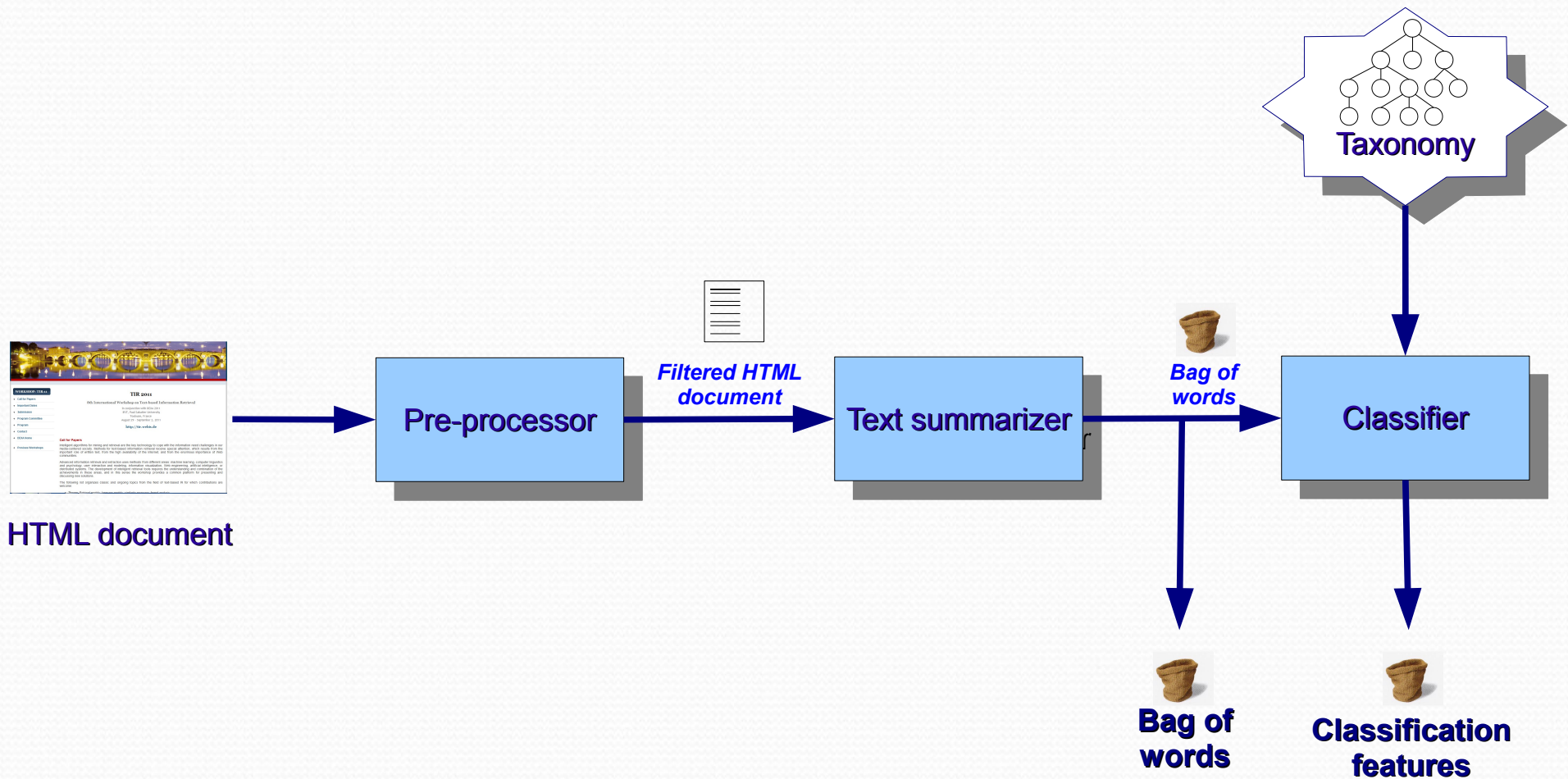


# TS: Experimental Results

	T	F2P	FLP	MK	MT	TFP	TFP	TF2P	TFLP	MTK
Precision	0.798	0.606	0.699	0.745	0.702	0.717	0.802	0.822	<b>0.832</b>	0.766
Recall	0.692	0.581	0.673	0.719	0.587	0.568	0.772	0.789	<b>0.801</b>	0.699
F-measure	0.729	0.593	0.686	0.732	0.639	0.634	0.787	0.805	<b>0.816</b>	0.731
Avg. Terms	3	13	24	24	25	15	16	27	26	34

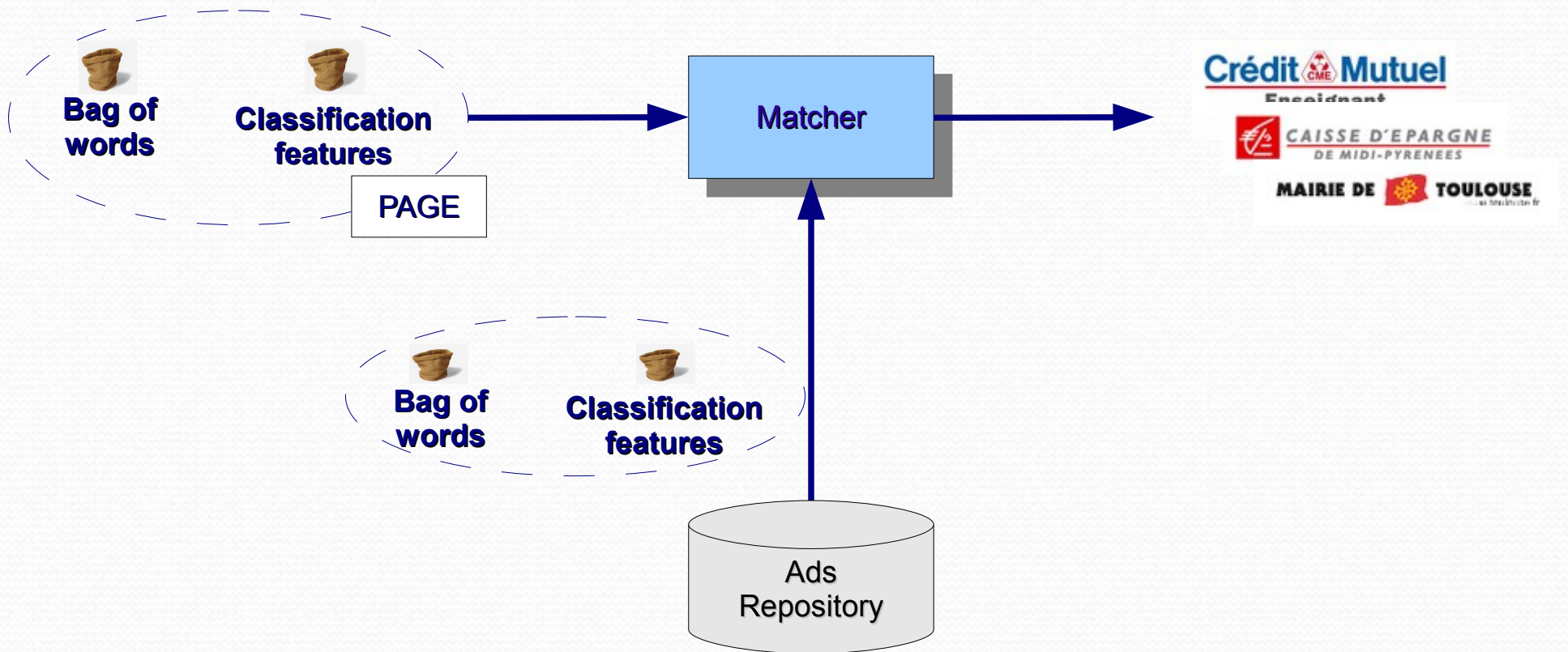
- Adding information about the title improves the performances
- Each novel method have better performances than classic methods
- The TFLP provides the best performance, as FLP does for the classic techniques

# CA: The Adopted System

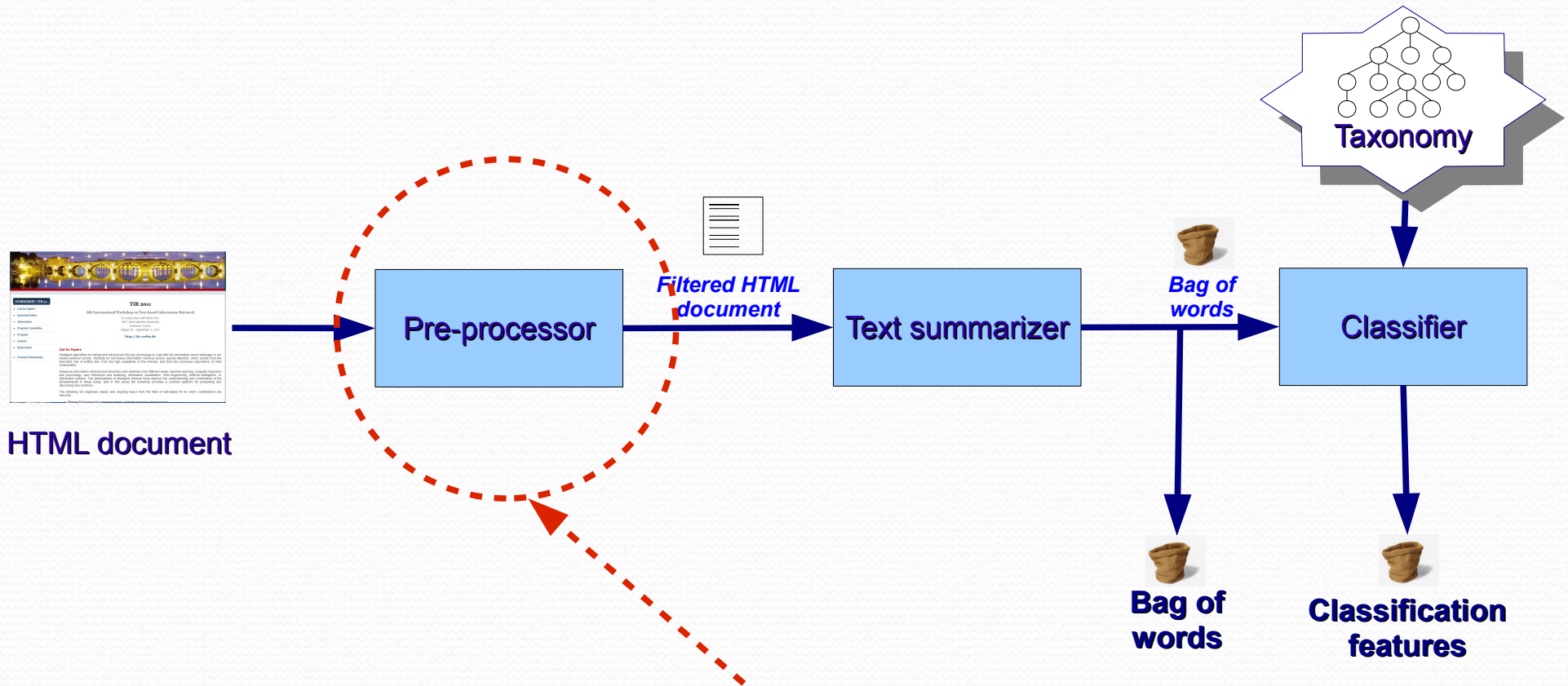




# CA: The Adopted System



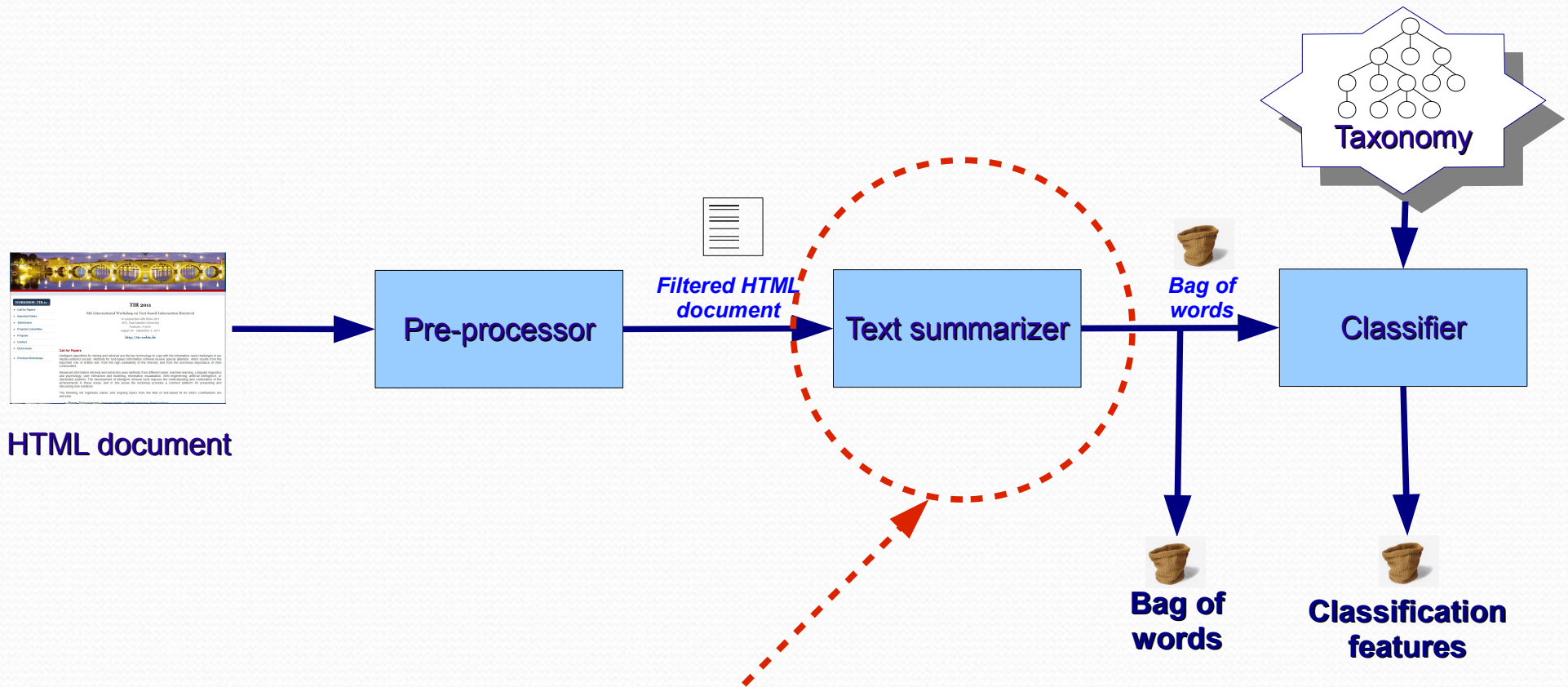
# Preprocessor



- Text extraction, stop-words removing, stemming



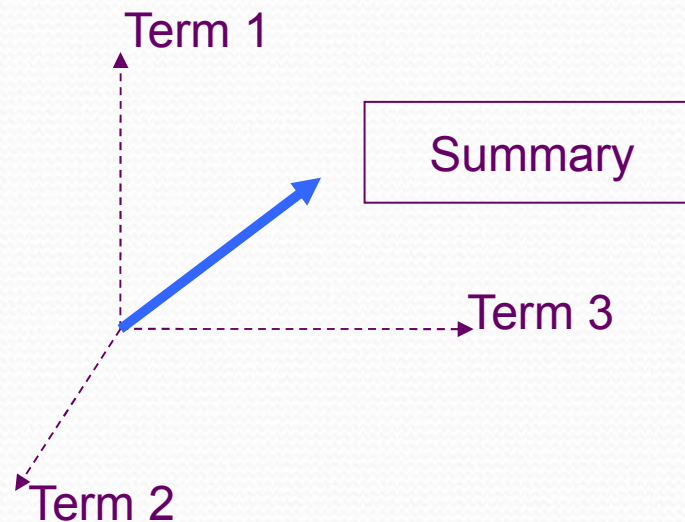
# Text Summarizer



- Summary of a web page

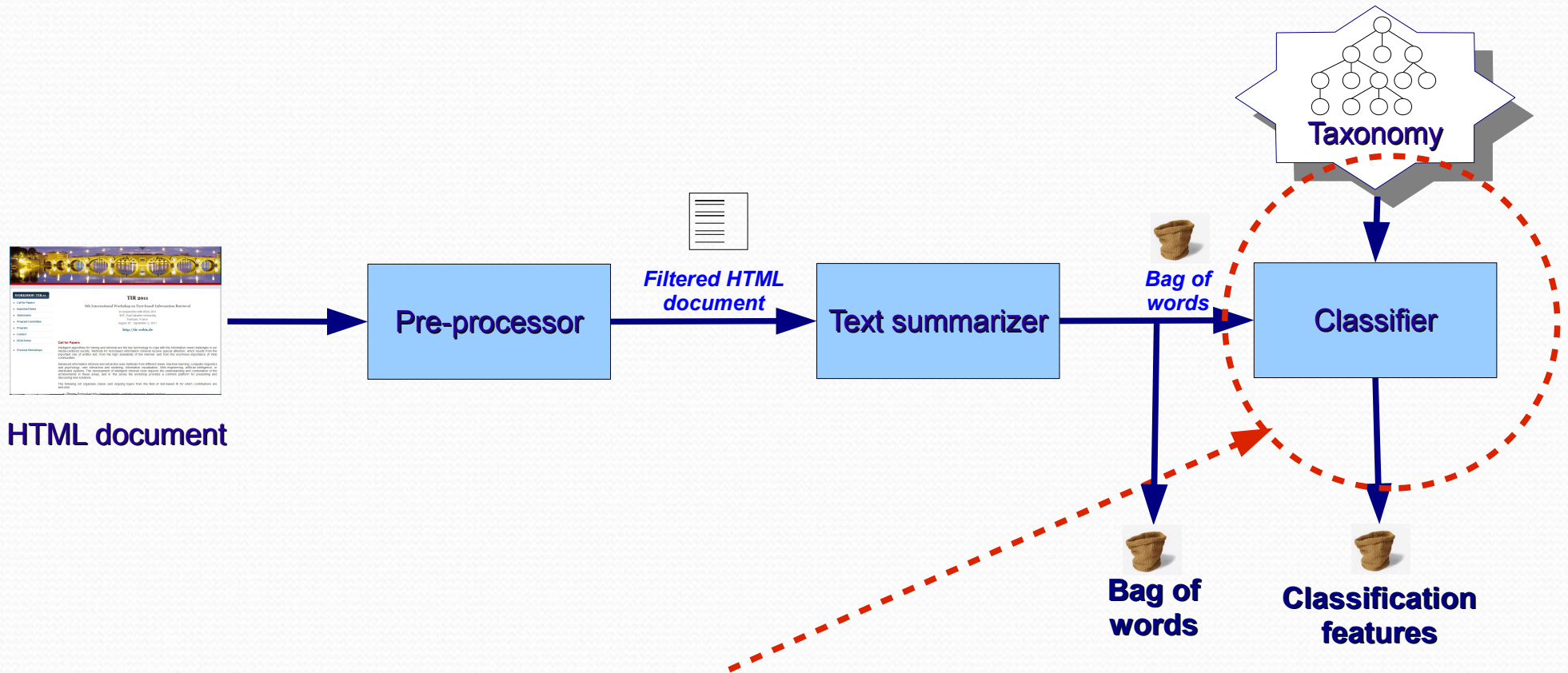
# Text Summarizer

- Implemented for each analysed technique
- Output
  - Vector-based representation of BoW (TF-IDF)





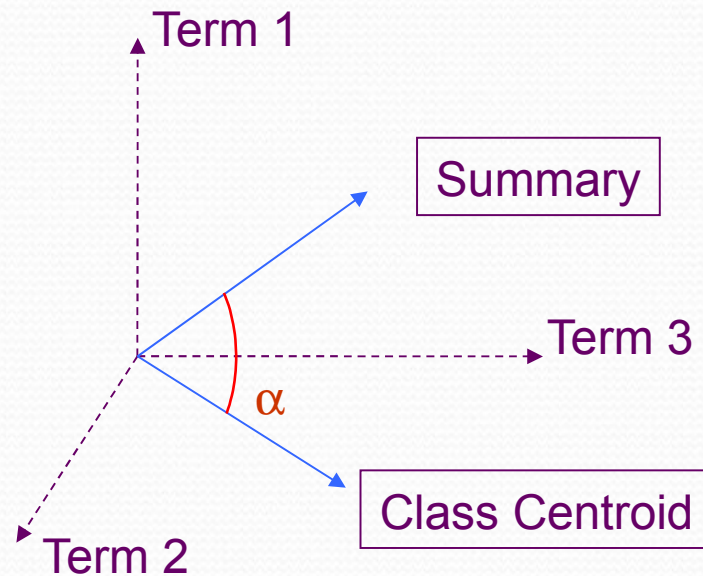
# Classifier



- Hierarchical classification

# Classifier

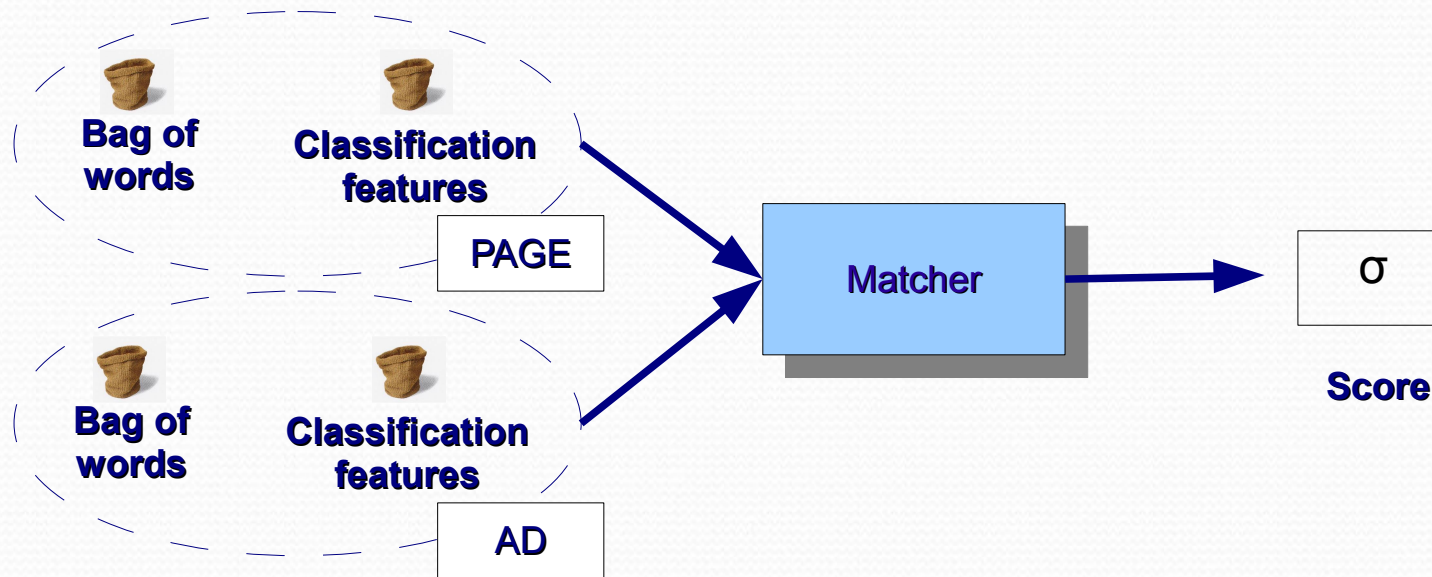
- Semantic centroid-based classification



- Output
  - Vector-based representation of CF of page and ads



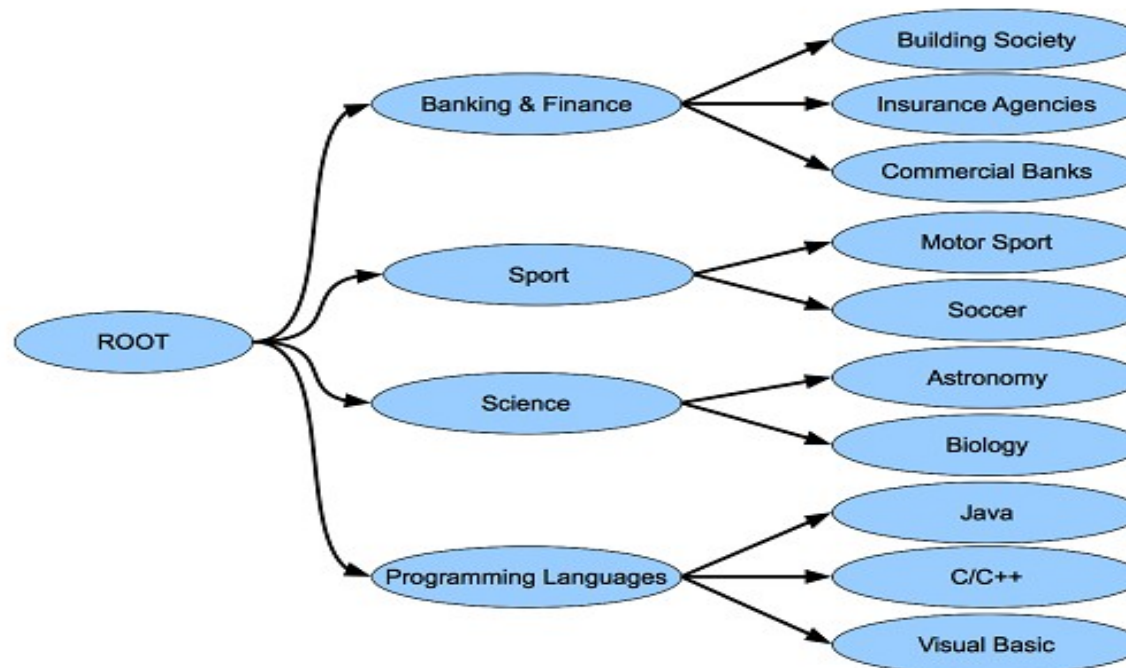
# Matcher



$$\sigma(P, A) = \alpha \cdot \text{sim}_{BoW}(P, A) + \beta \cdot \text{sim}_{CF}(P, A)$$

# The Impact of TS

- Adoption of each TS technique in the system and comparison
- Adopted Dataset (BankSearch)



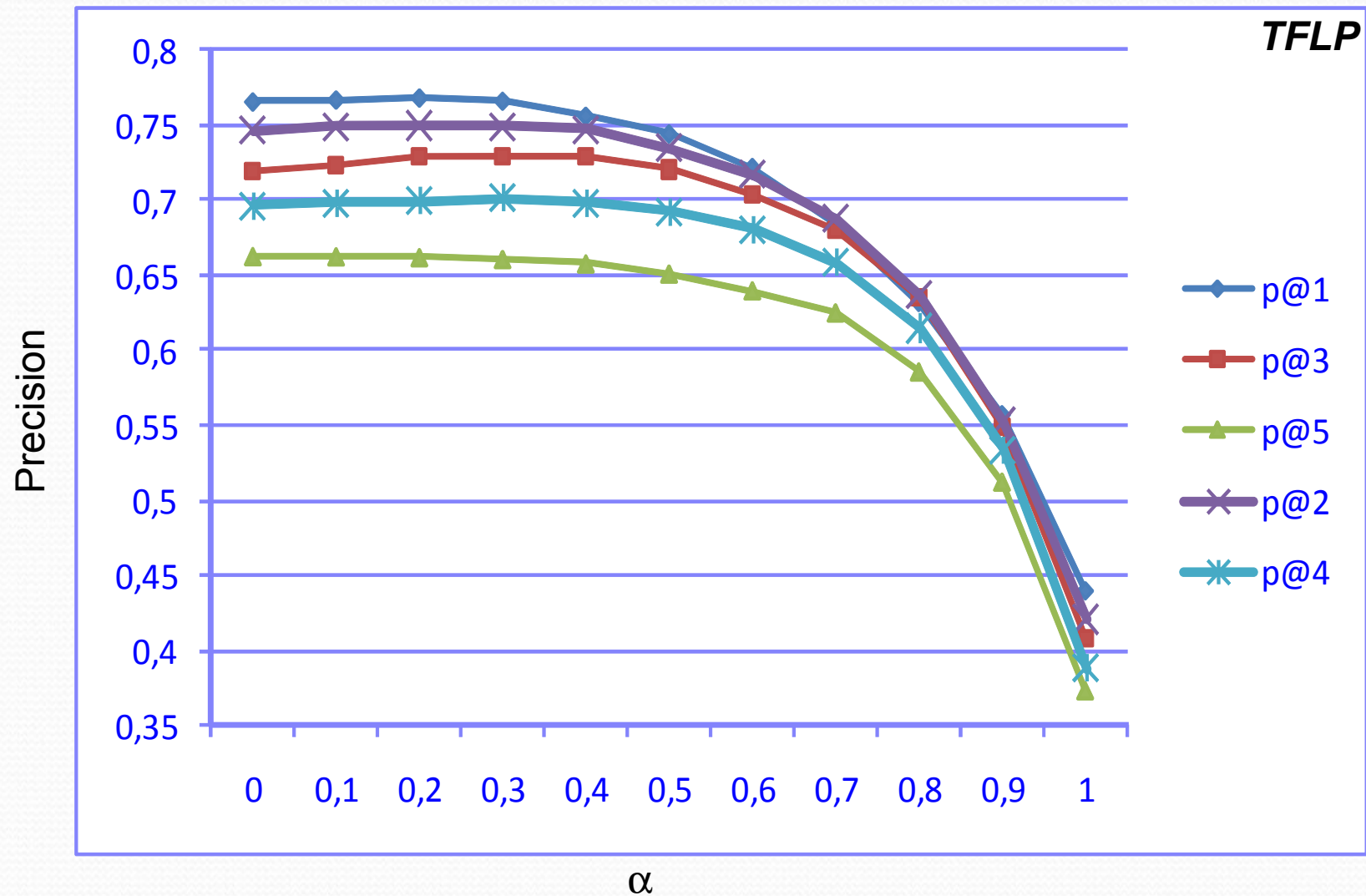


# Experimental results

	<i>p@1</i>	<i>p@2</i>	<i>p@3</i>	<i>p@4</i>	<i>p@5</i>
<i>T</i>	<b>0.680</b>	<b>0.675</b>	<b>0.652</b>	<b>0.639</b>	<b>0.595</b>
<i>FP</i>	0.488	0.476	0.448	0.421	0.391
<i>F2P</i>	0.613	0.609	0.588	0.558	0.514
<i>FLP</i>	0.674	0.653	0.617	0.582	0.546
<i>MK</i>	0.631	0.620	0.581	0.541	0.500
<i>MT</i>	0.640	0.617	0.610	0.586	0.547
<i>TFP</i>	0.744	0.707	0.691	0.669	0.637
<i>TF2P</i>	0.740	0.723	0.721	<b>0.712</b>	<b>0.678</b>
<i>TFLP</i>	<b>0.768</b>	<b>0.750</b>	<b>0.729</b>	0.701	0.663
<i>MTK</i>	0.711	0.698	0.685	0.663	0.608

- Title leads to an improvement of performances

# Experimental results





# Conclusions

- We presented a comparative study on TS applied to CA.
- We propose several techniques that improve the classic methods
- We evaluated the corresponding CA systems.
- Experimental results confirm the intuition that adopting TS techniques allows to improve performances in term of precision.
- As for future directions, further experiments are under way, with the adoption of larger dataset extracted by DMOZ.

Thanks for your  
attention!