

## Information Retrieval

INFO 4300 / CS 4300

---

- Classification applications in IR
  - Spam detection
  - Sentiment classification
- ➔ Online advertising
  - **Identifying fake online reviews**

## Advertising

---

- **Sponsored search** – advertising presented with search results
- **Contextual advertising** – advertising presented when browsing web pages
- Both involve finding the most relevant advertisements in a database
  - An advertisement usually consists of a short text description and a link to a web page describing the product or service in more detail

## Searching Advertisements

---

- Factors involved in ranking advertisements
  - similarity of ad text content to query/web page
  - bids for keywords in query
  - popularity of advertisement
- Small amount of text in advertisement
  - dealing with vocabulary mismatch is important
  - expansion techniques are effective

## Example Advertisements

---

**fish tanks at Target**

Find **fish tanks** Online. Shop & Save at Target.com Today.  
www.target.com

**Aquariums**

540+ Aquariums at Great Prices.  
fishbowls.pronto.com

**Freshwater Fish Species**

Everything you need to know to keep your setup clean and beautiful  
www.FishChannel.com

**Pet Supplies at Shop.com**

Shop millions of products and buy from our trusted merchants.  
shop.com

**Custom Fish Tanks**

Choose From 6,500+ Pet Supplies. Save On Custom **Fish Tanks!**  
shopzilla.com

Advertisements retrieved for query “fish tank”

## Searching Advertisements

---

- Expand query and/or document using the Web
- Employ pseudo-relevance feedback
  - rank exact matches first, followed by stem matches, followed by expansion matches
- Query reformulation based on search sessions
  - learn associations between words and phrases based on co-occurrence in search sessions

## Contextual ads

---

- Extract keywords from web page content and use to search ad DB
- Use ML techniques to determine which words and phrases on the web page are significant
  - tf, df, functions of tf and df
  - \*Frequency in query log
  - \*Location of word/phrase in the document
  - \*Capitalization, highlighting

## Another option: Classifying Online Ads According to a Semantic Hierarchy

---

- Unlike traditional search, online advertising goes beyond “topical relevance”
- A user searching for ‘tropical fish’ may also be interested in pet stores, local aquariums, or even scuba diving lessons
- These are semantically related, but not topically relevant!
- We can bridge the semantic gap by classifying ads and queries according to a semantic hierarchy

## Semantic Classification

---

- Semantic hierarchy
  - Example: Pets / Aquariums / Supplies
- Training data
  - Large number of ads and queries (with commercial intent) are manually classified into the hierarchy
- Nearest neighbor classification has been shown to be effective for this task
  - tf.idf cosine similarity measure
- Hierarchical structure of classes can be used to improve classification accuracy

# Semantic Classification

