

# Information Retrieval (IR): Between Big Data and Small Probabilities

Thomas Roelleke

# Between Big Data and Small Probabilities

---

- Traditional IR
  - Big data + simple and scalable methods and technology
  - Probabilities + IR models (some say “heuristics”)
- DB+IR
  - Technology for **complex information** management tasks
  - Example: From reviews, extract the facts and opinions about entities, and build a knowledge base for querying the data
- IR and Probability Theory
  - Web search tells: IR models work
  - Transfer **IR to pure probability theory, to other disciplines?**

# DB+IR

---

- No added sugar, No artificial colours, No genmod
  - What are the ingredients?
- NoSQL, NoJava, MiniPython, MiniPerl
  - High-level, data and knowledge-oriented languages
  - Rule-based, probabilistic reasoning over semi-structured data
  - For super-fast (hyper-agile) prototyping (if-then scenarios)

```
0.6 recommend(Product, Customer) :-  
    likes(Customer1, Product) &  
    similar(Customer, Customer1);  
0.4 recommend(Product, Customer) :-  
    customerProfile(Customer, Profile) &  
    productDescription(Product, Description) &  
    implies(Description, Profile);
```

# IR and Probability Theory

---

- IR: the important words are rare, very rare!
- Given an event that occurs once in 100 trials:
  - $P(x) = 1/100$
  - Examples:
    - a football defender scoring
    - a potential customer converting
- Probability to observe the event twice in only 10 trials?
  - In the world of traditional (Binomial) probabilities: relatively small
  - **Event occurrences, however, co-occur!**
  - Examples:
    - Ramos scores --- twice in one game!
    - A customer converts --- suddenly, there are several customers.

# Between Big Data and Small Probabilities

---

- DB+IR
  - Technology tailored to ***data analytics over semi-big data***
  - Super-fast, low-cost, transparent and hyper-agile prototyping
- IR and Probability Theory
  - ***Simple and effective math*** to rank candidates / make decisions
  - Complement existing probability theory (e.g. Bayesian)
- Applications
  - Senior Recruitment, Law Enforcement, Patent Search, NHS: information management tasks with a focus on difficult searches (rare events)