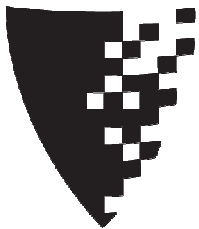


XML, DTD, and XML Schema

Introduction to Databases

CompSci 316 Fall 2014



DUKE
COMPUTER SCIENCE

Announcements (Tue. Oct. 21)

- **Midterm scores and sample solution** posted
 - You may pick up graded exams outside my office
 - Mean: 83.9
 - Stdev: 11.0
 - Max: 100+5
- PHP and Django example website code posted; more to come
- **Homework #3** to be assigned on Thursday
- **Project milestone #1** feedback to be returned this weekend

```
95 *****
90 *****
85 *****
80 *****
75 *****
70 ****
65 *****
60 **
```

Structured vs. unstructured data

- Relational databases are highly structured
 - All data resides in tables
 - You must define schema before entering any data
 - Every row conforms to the table schema
 - Changing the schema is hard and may break many things
- Texts are highly unstructured
 - Data is free-form
 - There is no pre-defined schema, and it's hard to define any schema
 - Readers need to infer structures and meanings

What's in between these two extremes?



Jun Yang

Professor
[Department of Computer Science](#)
[Duke University](#)

[Home](#)
[Publications](#)
[Students](#)
[Teaching](#)
[Personal](#)

Published Work

1. You Wu, Pankaj K. Agarwal, Chengkai Li, Jun Yang, and Cong Yu. "Toward computational fact-checking." *Proceedings of the VLDB Endowment*, 7(7):589-600, 2014. [[paper](#)]
2. Naeemul Hassan, Afroza Sultana, and Jun Yang. "FactWatcher." *Proceedings of the VLDB Endowment*, 7(7):589-600, 2014. [[paper](#)]
3. You Wu, Brett Walenz, Peggine Li, and Jun Yang. "Lies, damned lies, and statistics: A demonstration track." *Proceedings of the VLDB Endowment*, 7(7):589-600, 2014. [[paper](#)]
4. Albert Yu, Pankaj K. Agarwal, and Jun Yang. "Chicago, Illinois, USA, March 2014." *Proceedings of the VLDB Endowment*, 7(7):589-600, 2014. [[report](#)]
5. Afroza Sultana, Naeemul Hassan, and Jun Yang. "International Conference on Data Engineering." *Proceedings of the VLDB Endowment*, 7(7):589-600, 2014. [[report](#)]
6. Risi Thonangi and Jun Yang. "International Conference on Data Engineering." *Proceedings of the VLDB Endowment*, 7(7):589-600, 2014. [[report](#)]
7. Botong Huang, Shivnath Balakrishnan, and Jun Yang. "International Conference on Data Engineering." *Proceedings of the VLDB Endowment*, 7(7):589-600, 2014. [[report](#)]
8. Yi Zhang, Kristian Lum, and Jun Yang. *Engineering*, 25(5):1042-1049, 2014.
9. Pankaj K. Agarwal, Lars Argemann, and Jun Yang. *Geometry: Theory and Applications*, 2014.
10. Albert Yu, Pankaj K. Agarwal, and Jun Yang. *Engineering*, 24(10):1833-1840, 2014.

www.amazon.com/s/ref=sr_nr_n_6?rh=n%3A2625373011%2Cp_n_theme_browse-bin%3A2650365011%2Ck%3Asimpsons&keywords=simpsons&ie=UTF8&qid=1413...

Sponsored by Hershey's **Halloween Shop** [Shop now](#)

Shop by Department Movies & TV Hello, Jun [Your Account](#) [Try Prime](#) [Cart](#) [Wish List](#)

Movies & TV New Releases Best Sellers Deals Blu-ray TV Shows Kids & Family Anime All Genres Amazon Instant Video Prime Instant Video Your Video Library Trade-In

1-16 of 167 results for Movies & TV : Kids & Family : "simpsons" Sort by [Relevance](#)

Show results for [Any Department](#)

Movies & TV
 TV (70)
 Movies (34)

Refine by

Eligible for Free Shipping
 Free Shipping by Amazon

Format

DVD (110)
 Blu-ray (4)
 Amazon Instant Video (37)
 VHS (15)
 + See more


Genre

[Any Genre](#)
 Comedy
Kids & Family
 Action & Adventure
 Drama
 Mystery & Thrillers
 Westerns
 Romance
 Documentary
 Horror

Format


[See more...](#)

The Simpsons Movie 2007 [PG-13](#) [CC](#)

 [Shop Instant Video](#)
\$9.99 - \$12.99
 Buy movie
 Watch **instantly** on your PS3, Xbox, Kindle Fire, iPad, PC and other devices.


★★★★☆ (395)
 Runtime: 1 hr 27 mins
 Starring: [Dan Castellaneta, Julie Kavner, et al.](#)
 Directed by: [20TH_CENTURY_FOX](#)

The Simpsons Season 1 1989 [CC](#)

 [Shop Instant Video](#)
\$1.99 - \$17.99
 Buy episodes or Buy season
 Watch **instantly** on your PS3, Xbox, Kindle Fire, iPad, PC and other devices.

★★★★★ (701)

The Simpsons Season 26 2014 [CC](#)

 [Shop Instant Video](#)
\$1.99 - \$44.99
 Buy episodes or Buy TV Season Pass
 Watch **instantly** on your PS3, Xbox, Kindle Fire, iPad, PC and other devices.

★★★★★ (1)

Semi-structured data

- Observation: most data have some structure, e.g.:
 - Book: chapters, sections, titles, paragraphs, references, index, etc.
 - Item for sale: name, picture, price (range), ratings, promotions, etc.
 - Web page: HTML
- Ideas:
 - Ensure data is “well-formatted”
 - If needed, ensure data is also “well-structured”
 - But make it easy to define and extend this structure
 - Make data “self-describing”

[Click here to download full PDF material](#)