Wolfram|Alpha

- Answering Questions with the World’s Factual Data

Joshua Martell - Feb 2nd @ 5pm, Mission City B5
What is Wolfram|Alpha?

- Computational Knowledge Engine
  - the world’s systematic knowledge computable and accessible to everyone
  - to compute whatever can be computed about anything
  - Integrate the world’s facts, data, and algorithms
  - 5 years of R&D
  - launched in May 2009
Wolfram|Alpha by example

- when is valentines day?
- how long was t-rex?
- how many calories is in a burger?
Examples: Math

- what is 2+2?
- plot sin(x^2+y^2)
- int e^t sin(5t) dt
- product (1-1/n^4), n=2 to infinity
Examples: Data lookups

- population of china
- GDP of the EU
- flight time from seattle to tokyo
- AAPL MSFT
Examples: Visualization

- earthquakes dec 2004
- qr code: http://strataconf.com
- caffeine
Examples: Formulas

- mortgage 5% 20 yr
- RLC circuit
Examples: Fun

- scrabble quixotic
- what's the meaning of life?
- airspeed velocity of an unladen swallow
Elements of the output

Described in the tour:
- input field
- assumptions (sometimes)
- input interpretation
- results pod
- other pods
- buttons and pull-down menus
- output extras under the pod
- source information
More about the input interpretation

- getting from the free form input to here is our secret sauce
- combination of heuristics, algorithms, and developer curation
- W|A has formed an exact expression representing your input
More about pods

- W|A looks for components that can report about the input
- results pod is the “answer”
- pods load asynchronously
- related cross-domain information in other pods
- entity only inputs
Finding data

- want to have the best information available
- quality, breadth, and technical assessment of each source
- web searches, evaluation of source materials
- prefer primary sources, prefer digital sources
Finding data

- technical considerations: Print >> PDF >> HTML >> DB >> CSV
- discuss with world experts to understand the data
- PD & US Gov have simple licensing terms
- corporate deals are more complicated, but data is better quality, documentation, assistance
Aggregating data

- rare that one source has it all
- fill in with secondary sources
- alignment is difficult, troublesome, and error prone
- use common identifiers, verify data across sources
- hand checking....does this value make sense?
- automating updates
Data cleaning and curation

- we want the best data, and we’re willing to work for it
- automate as much as possible, do the rest by hand
- use Mathematica to find outliers and oddballs, explore data, verify quality
- takes time and attention to detail
Making it computable

- teach W|A about the domain and its relationships to existing domains
- natural language parsing for entities and properties
- data becomes a building block for the inspiration of users
- GDP Greece / population of Italy
- questions thus far?
Data storage and retrieval

- read heavy system
- writes from feeds and developers
- some results involve only computation, but usually some data is used
- elaborate tracking of data changes
Versioning and deployment

- significant development effort
- different tools for programmatic changes verses hand curation
- versioning is closely tied to deployment; deploy only new or updated values
- weekly deployment of a new revision of W|A with data updates and code changes
- content distribution system updates colocations
Purely computational data

- formulas, encodings, etc.
- internal APIs are very flexible
- data comes from a wide variety of computational and unconventional sources
- many, many built in algorithms in *Mathematica*
Computation and visualization

- *Mathematica* as a development platform
- our (not so) secret weapon
- functional, very high level, symbolic programming language
- built in everything
- statistics, numerics, advanced plots, charts, file formats
- large collection of algorithms
- database/Java/.NET integration - C interface
- used by
  - all 15 major US Federal government departments
  - all Fortune 50 companies
  - all 50 largest universities worldwide
Web components

- web Mathematica = Mathematica powered web pages
- Mathematica integrates into the servlet engine
- majority of code is Mathematica
**Other Technologies: API**

- REST API, various language bindings
- returns XML encoded HTML, plaintext, images, etc
- used for W/A iPhone iPad / Android app
- used by Bing to display W/A results into search results
- free to try out

10363 seconds
Other Technologies: Wolfram|Alpha appliance

- W|A in your data center
- consulting to integrate corporate data into W|A system
Data Summit 2011

- Wolfram Data Summit - Washington DC, Sep 7,8, 9
- not W|A specific
- many corporate, non-profit, government attendees
- meeting place to discuss common issues and solutions