SCALABLE WEB PROGRAMMING

CS193S - Jan Jannink - 3/11/10

WOW

* Terrific energy at the demo lunch yesterday
* Sustained volume of discussion was impressive
* Feedback from guests and students was great
* Let me know if I can post links to your projects

Things I'd Improve

* Get the class on Google Wave
    * some of you weren't catching the lectures in a timely fashion
* Set up git and continuous integration in week 1
* Have one additional checkpoint and start the projects sooner
* Invite a couple of guest speakers who also think about scalability
* Find smarter ways to integrate remote students into projects

Things I Felt Worked

* Ending up with 6 demo worthy projects
* Self assembling teams of about 4-8 people
    * learning first hand about team interaction problems
* Building, testing and brainstorming about your code
* Having a project motivation that was not grade related

Things I Hope You Take Away

* Constraint management is an engineerable art
  * maximizing results given your limitations
* Communication is the biggest key to human scalability
* IT tools for SW development are also for communication
* All of this is more broadly applicable than just to web apps

Tell Me What You Think

Future Web App Technology

* Web data structures
    * new tribes
* Algorithms
    * web computation
* Odds and ends and Speculation

Five Random Things

* Noticed CA license plate 2TOT003
* Daughter spoke about a future event (mother going off to work)
* Talked about seven deadly sins and the consumer web
* Pondered how Sesame Street might use -1 as number of the day
* Saw ducks, thought about ducktyping for today's lecture
    * "looks like a duck, quacks like a duck"

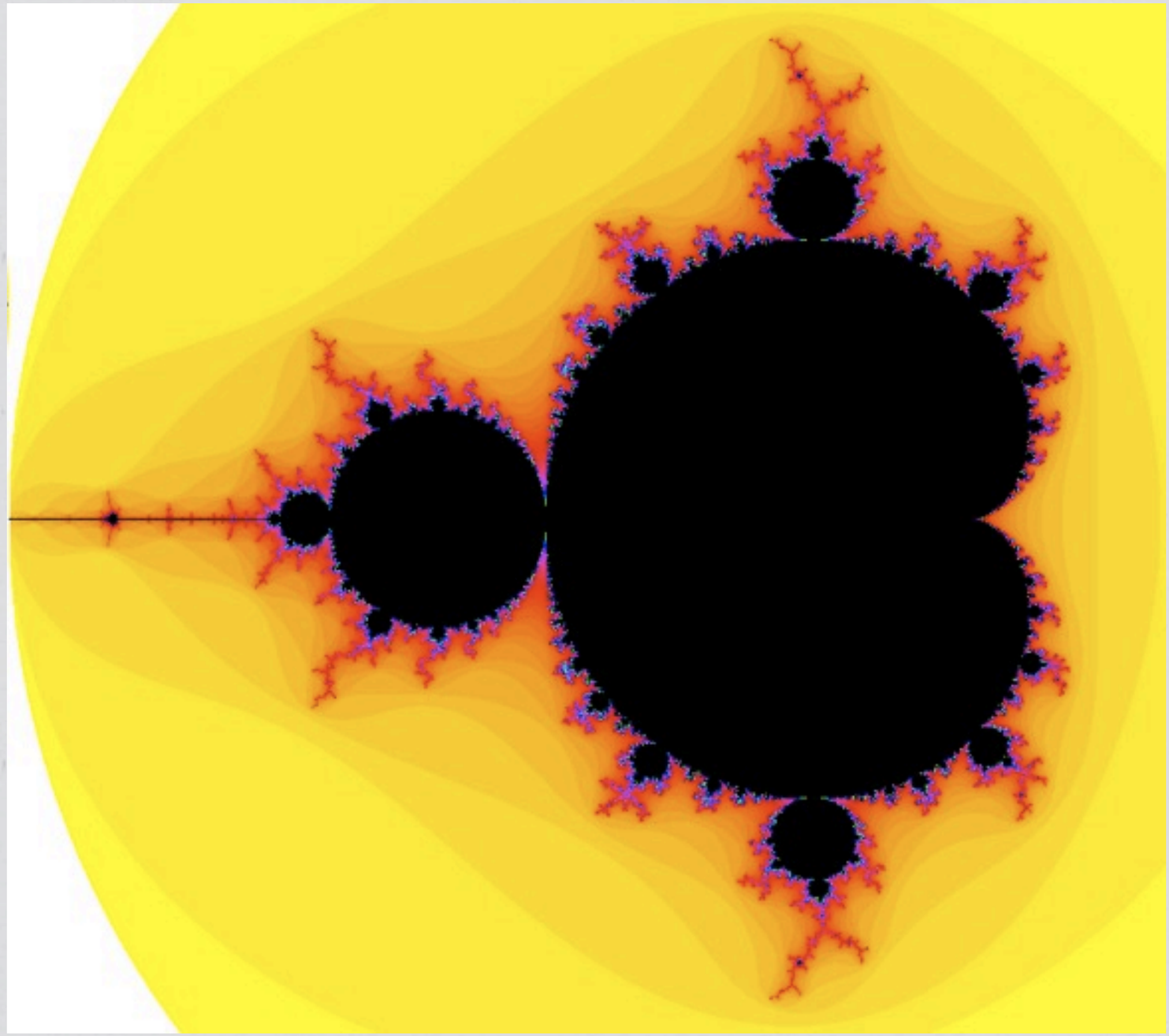
99% of Everything is Crap

* Jan's (optimistic) corollary
    * 10% of that crap is fertilizer
* Jan's (wasteful) lemma
    * non crap can't emerge without 100 times as much crap
* Jan's Augean (scatological, coprological) task
    * observe all the crap to see if/when non crap emerges

Ahem, Relevance Please?

* Managing the world's information
* Exploiting the (key, value) pair model for data management
* Assuming the web (graph) model of information is universal
* Recognizing that most data will be unstructured

Complex Frontiers: $Z = Z^2 + C$



The New Tribes

* Historically we have always pigeonholed people
    * kinship, geography, language, beliefs, appearance, gender
* The web is fundamentally altering this as we speak
    * online our tribes are defined by tagclouds
* Tagclouds are fuzzy graph approximations of people, concepts
    * enough overlap indicates group membership

Already in Use

* Music affinity is an expressive shorthand for the broader tagcloud
  * Napster, imeem exploited this to great advantage
* Duck typing is emblematic of this in programming languages
  * drawback: static type checking incomplete at best
  * define persistence model to support flexible types

Future Applications

* Generalized data matching, approximation algorithms
* Mental models
    * we really are only a sum of serendipitous events
    * some stick, some don’t
        * how many of my five will you remember tomorrow?

Algorithms

* PageRank variants
* Efficient graph instantiation
  * when it doesn’t fit in memory
* Relevant subgraph extraction
  * use a search term or graph as input
* Web summarization

Thinking Big

* Over 6 billion lines of open source software
    * biggest software “company”
    * Biggest are linux kernel, eclipse, java
* Over 250,000 projects
    * explosion of computer language and database projects
* Best option for implementing mature technologies

Future Web App Tech

* PHP beats GWT for rapid prototyping, GWT more maintainable
    * the speed of code-[compile]-test loop factors into that a lot
        * is this a permanent tradeoff or can we achieve both?
            * embed java in html?
            * can we make browser jvm work like javascript interpreter?
* Will Eclipse ever be fast as an IDE?

Worth Checking Out

* Open source code search engines
  * http://www.koders.com/
  * http://www.google.com/codesearch

Q & A Topics

* Is it possible to teach scalability in one quarter?
* Predictability horizon has significantly shortened
    * can we speak with confidence about the web in 2015?
* Radical originality/innovation is more difficult to achieve
    * the same concepts are broadly available to many people

Final Thoughts

* This course dedicated to a mentor, Rajeev Motwani