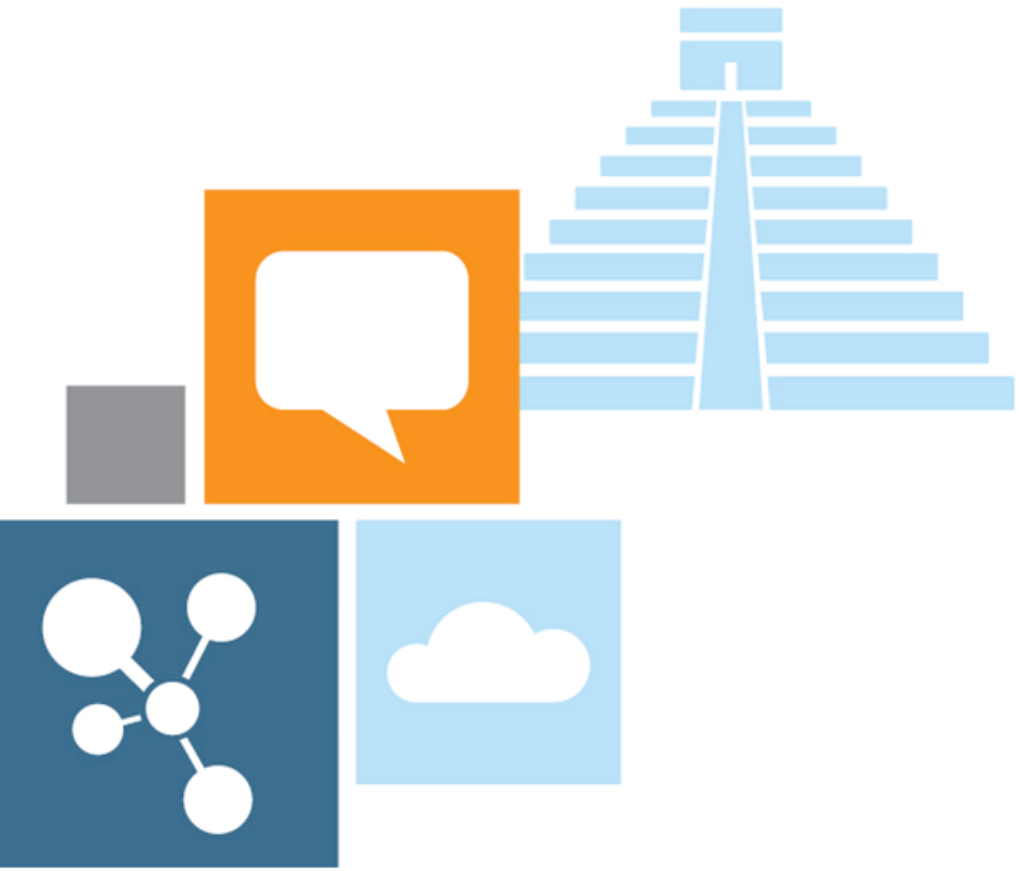
Microsoft



Microsoft® Research

Faculty Summit 2012

Riviera Maya, Mexico | May 23-25 | In partnership with CONACYT

Why Software Engineering is the Best Job in 2012

Judith Bishop
Director of Computer Science
Microsoft Research Connections

From a report in March 2012

Based on

- Work Environment
- Physical Demands
- Job Outlook
- Income Levels
- Stress

http://www.careercast.com/jobs-rated/10-best-jobs-2012

Some History

Eclipse software development environment with its extensible plug-in system.

- From IBM Canada, 2001
- Now in a Foundation
- Free and open source under its own license
- Strong community ba

Java programming language and its run-time platform

- From Sun Microsystems in 1995
- Now with Oracle
- Free and Open Source under GNU public license
- Strong research, education and developer base
- Part of browser technology

Microsoft collaborates with universities

- Why?
  - Expand a lab's research base from 100s to 10,000s
  - Verify the work in practice
  - Launch new applications
  - Support tomorrow's leaders
- How?
  - Awards, grants, prizes
  - Interns (1,000 a year)
  - Visitors
  - And more

Amazing numbers

- More than 100 top PhD research students from leading European universities in Microsoft Research Fellowships
- More than 25 000 scientists, academics and students have attended Microsoft-sponsored conferences, summits and workshops since 2005
- Nearly 1 000 interns selected for internships at our labs in Redmond, China and India annually
- 465 students from more than 35 countries have interned at the Cambridge lab in the past 7 years
- 50 conferences sponsored in the past 5 years for $350,000

Computer Science Activities

- Innovation Foundation (SEIF)
- Interaction with RiSE
- Promoting Visual Studio through ICSE events

Software Engineering

- Interaction with NLP, ISRC, Bing
- Developed Language Models Services
- Bing-MSR Speller Challenge
- Developing Knowledge Services

Semantic Computing

- Interaction with Network Research team in MSR-R
- Student projects on WP7 phones at 17 universities
- Promoting Azure and WP7

Mobile Computing

- Book on Patterns and Practices group
- Course on Concurrency and Parallelism Promoting C# and F#

Parallelism and Concurrency

- Developing a Try F# system in a browser
- Workshops on "F# in Education"
- Supporting Pex4Fun
- TouchDevelop outreach and book

Programming Languages

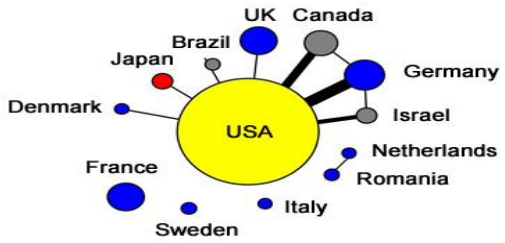
- CS EdWeek
- Work with the NSF, ACM, IEEE and IFIP
- MSR Software Summits
- Support for 12 conferences

Community interaction

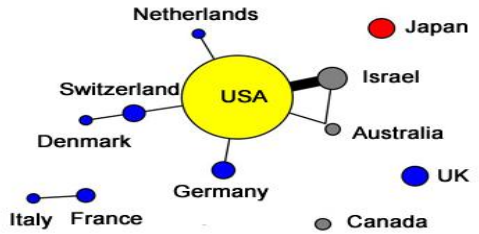
CS is research-oriented with strong internal and external impact

How collaboration has grown, 1996-2011

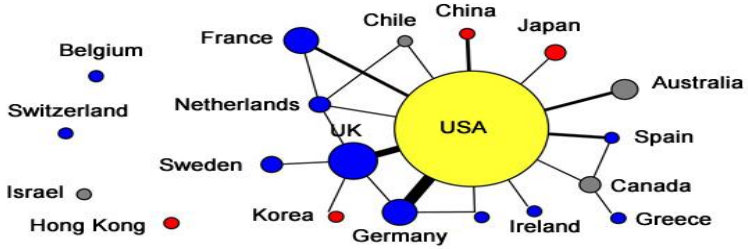
Mapping Collaboration Networks in 'Programming' Conferences



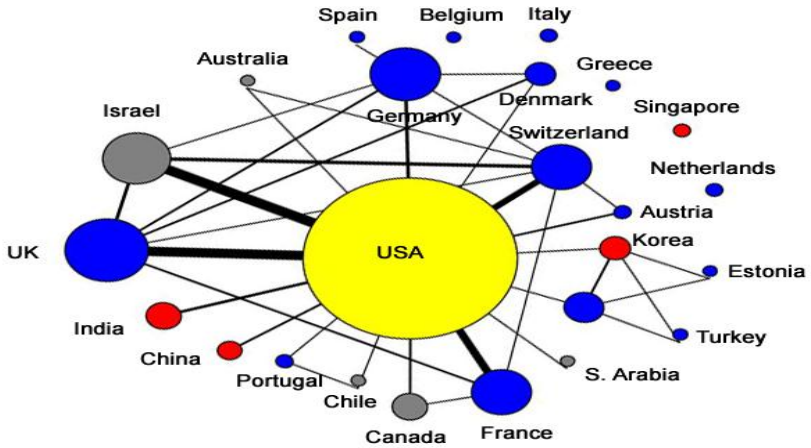
1996



2001



2006

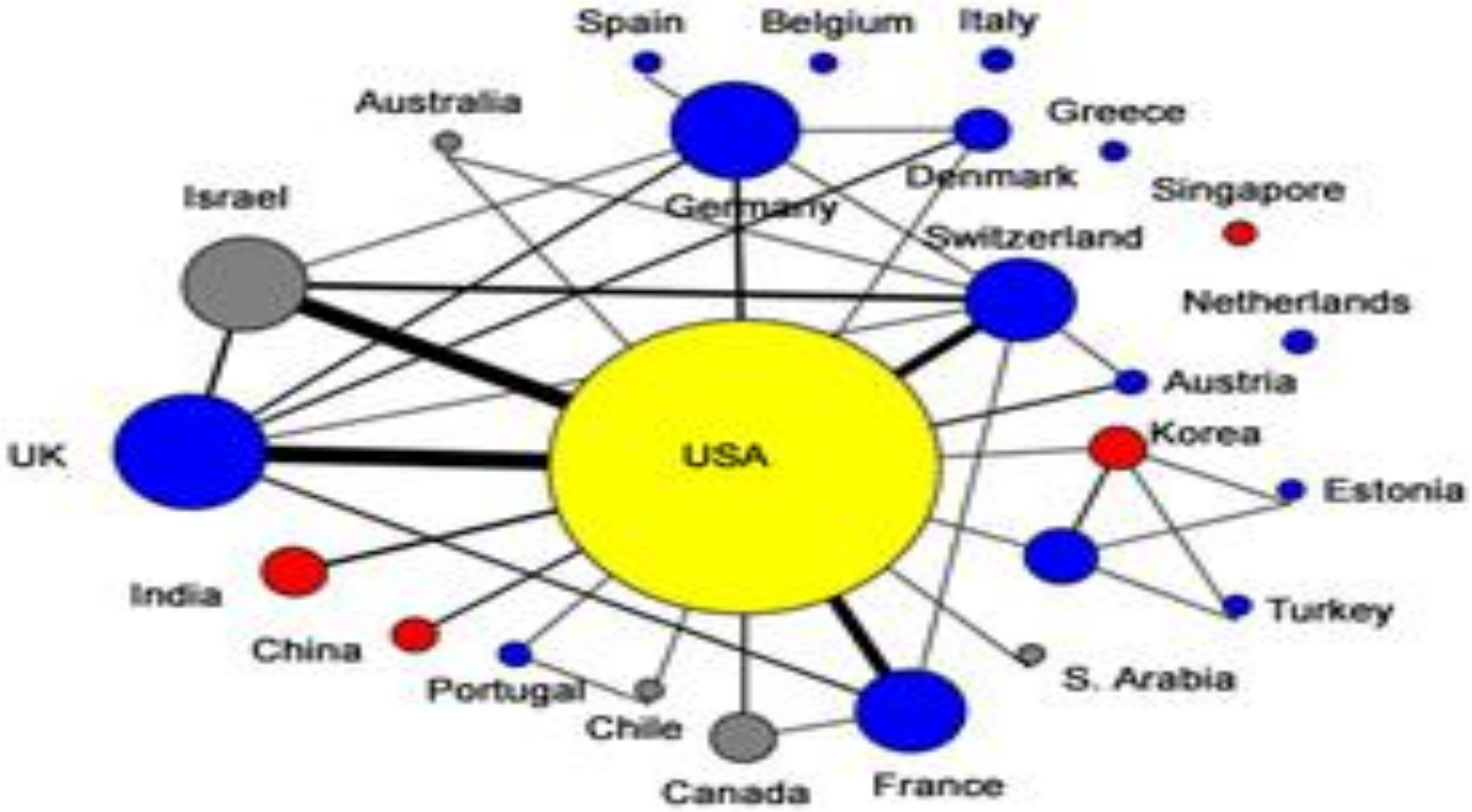


2011

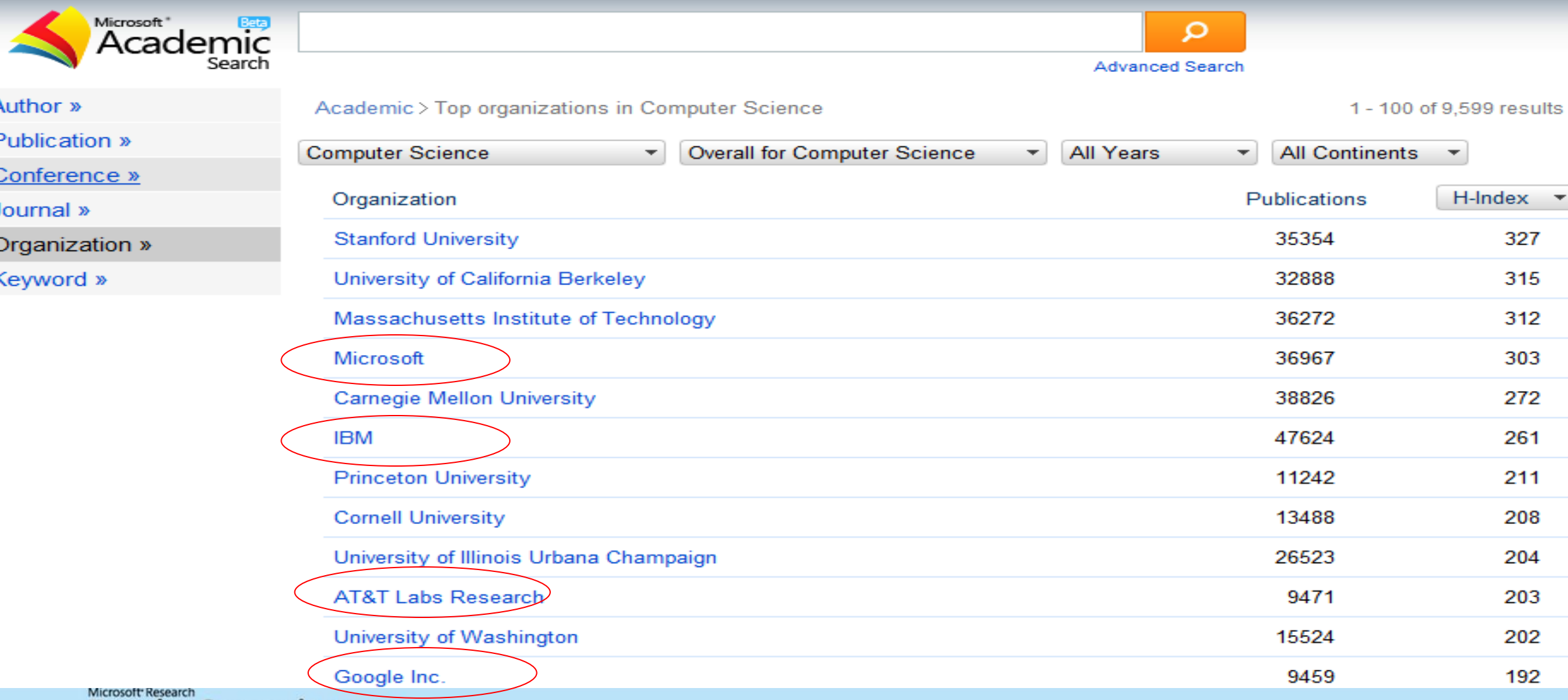
*** Circle size represents number of publications authored by each country. Lines between countries indicate publication co-authorship. Line width indicate the degree of collaboration (wider = more collaborations). Conferences analyzed include PPOPP, POPL, PLDI, and OOPSLA.**

Microsoft

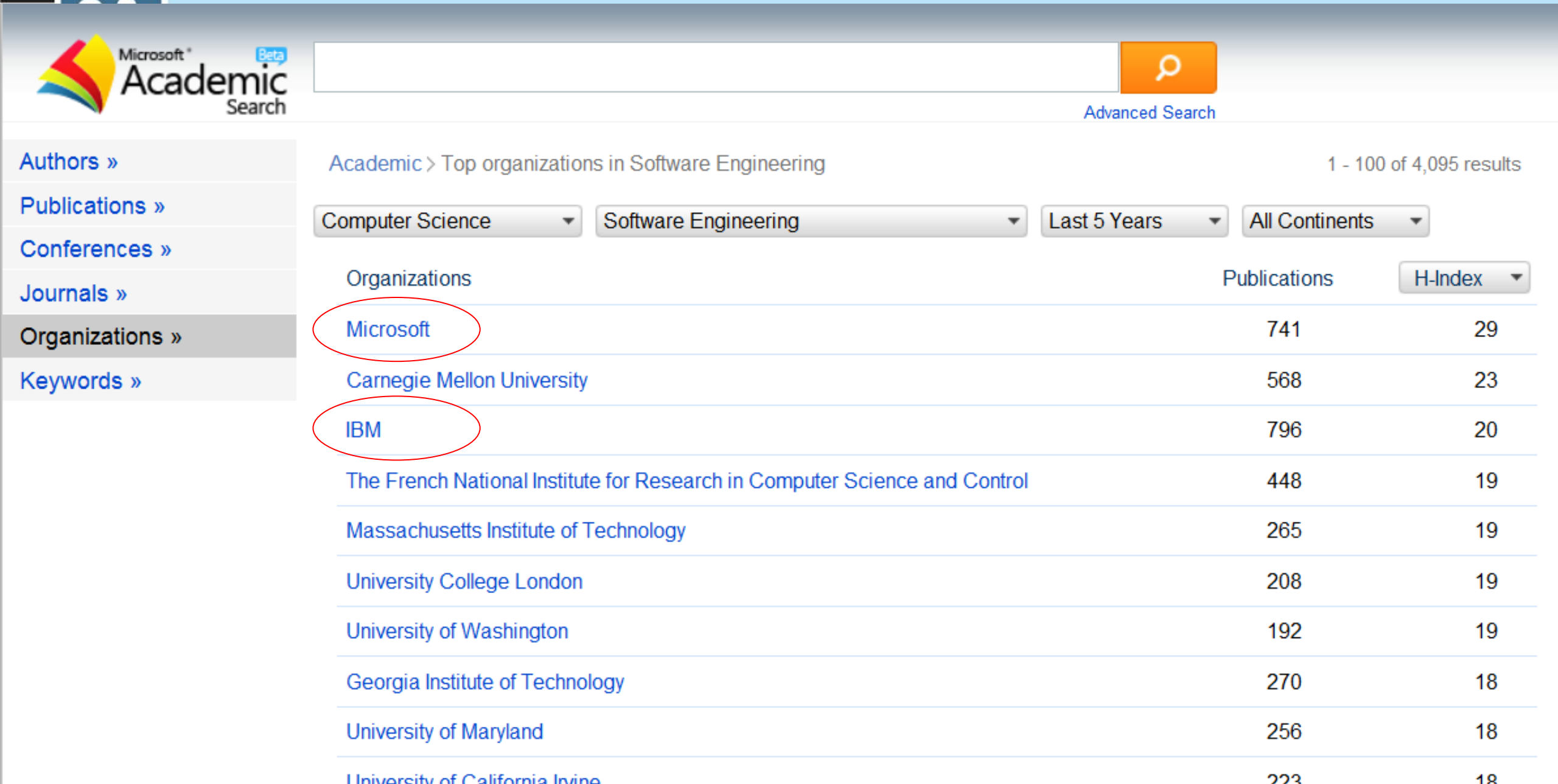
PL Collaboration 2011



Top 12 CS Research Organizations



Top Software Engineering organizations last 5 years



Agenda

1. Software engineering in the large
2. Software in the browser
3. Software social experience
4. Social aspects in the large
5. Software development kits

2. SOFTWARE ENGINEERING IN THE LARGE

The Challenge

Microsoft ships software to 1 billion users around the world

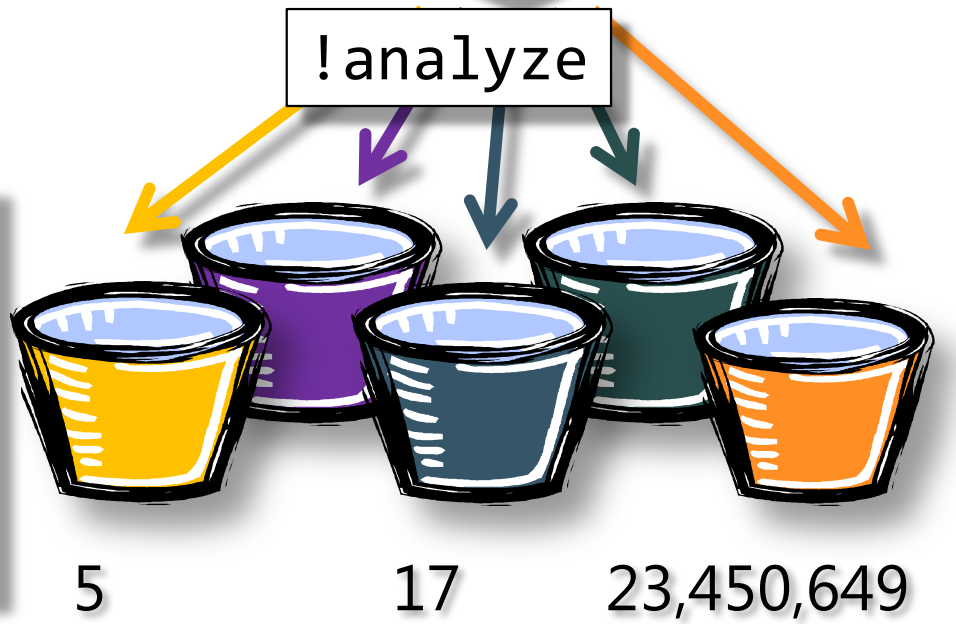
- How do we find out when things go wrong?

We want to

- fix bugs **regardless of source**
- prioritize bugs that **affect most users**
- **generalize the solution** for any programmer
- get the **solutions out to users** efficiently
- try to **prevent bugs** in the first place

Debugging in the Large with WER…

Minidump



5 17 23,450,649

!analyze

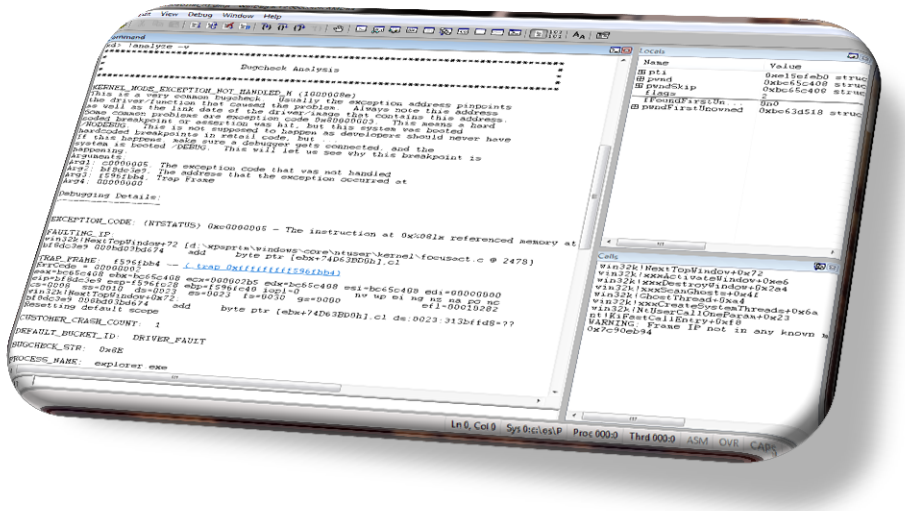
Engine for WER bucketing heuristics

Extension to the *Debugging Tools for Windows*

- input is a minidump, output is bucket ID
- runs on WER servers (and programmers desktops)

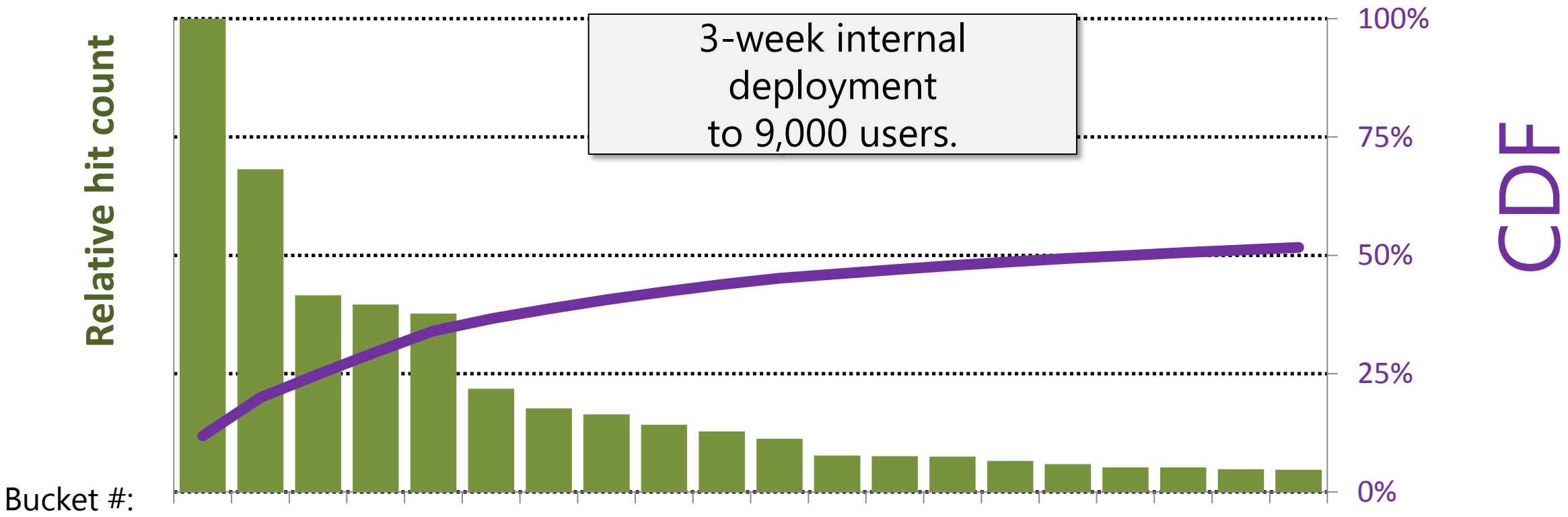
500 heuristics

- grows ~ 1 heuristic/week



Kirk Glerum, Kinshuman Kinshumann, Steve Greenberg, Gabriel Aul, Vince Orgovan, Greg Nichols, David Grant, Gretchen Loihle, and Galen Hunt, Debugging in the (Very) Large: Ten Years of Implementation and Experience, in *Proceedings of the 22nd ACM Symposium on Operating Systems Principles (SOSP '09)*, Association for Computing Machinery, Inc., Big Sky, MT, October 2009

Top 20 Buckets for *MS Word 2010*



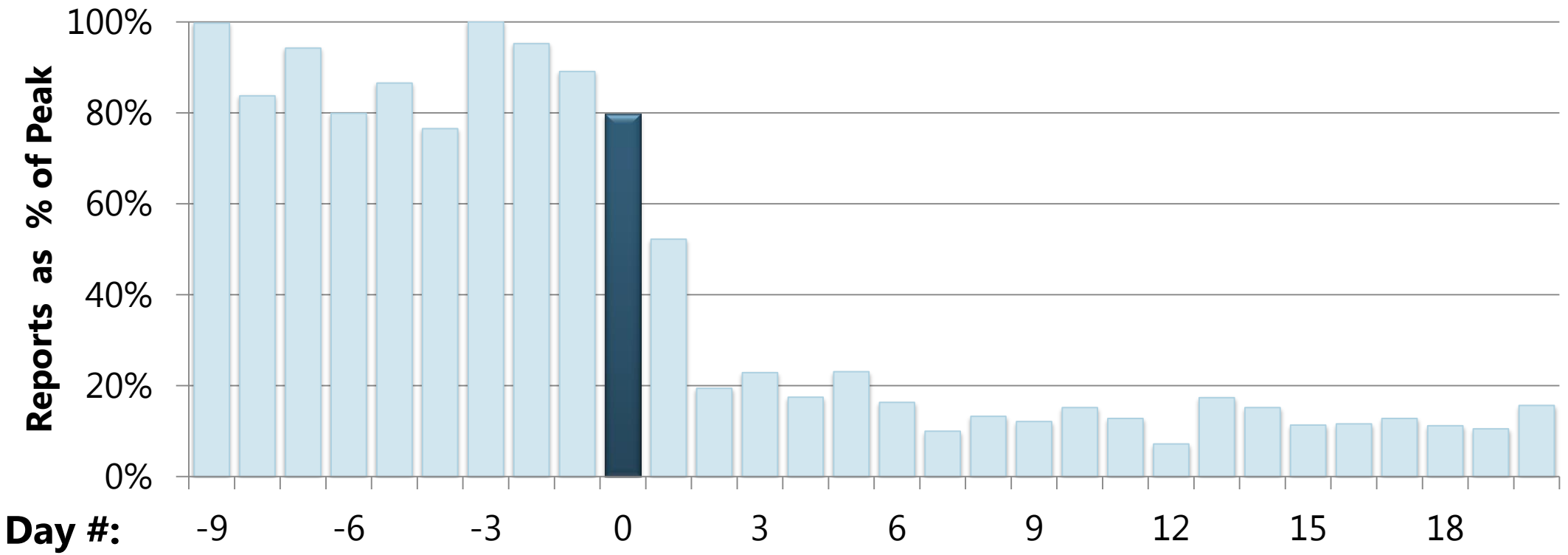
Just 20 buckets account for **50% of all errors**

Fixing a small # of bugs will help **many users**

Fixing bugs in software

- First use found >=**5-year old** heisenbugs in Windows
- Windows Vista team **fixed 5,000 bugs** in beta
- Anti-Virus vendor fixed top 20 buckets and **dropped from 7.6% to 3.6%** of all kernel crashes
- Office 2010 team fixed 22% of reports **in 3 weeks**
- And you can fix yours…

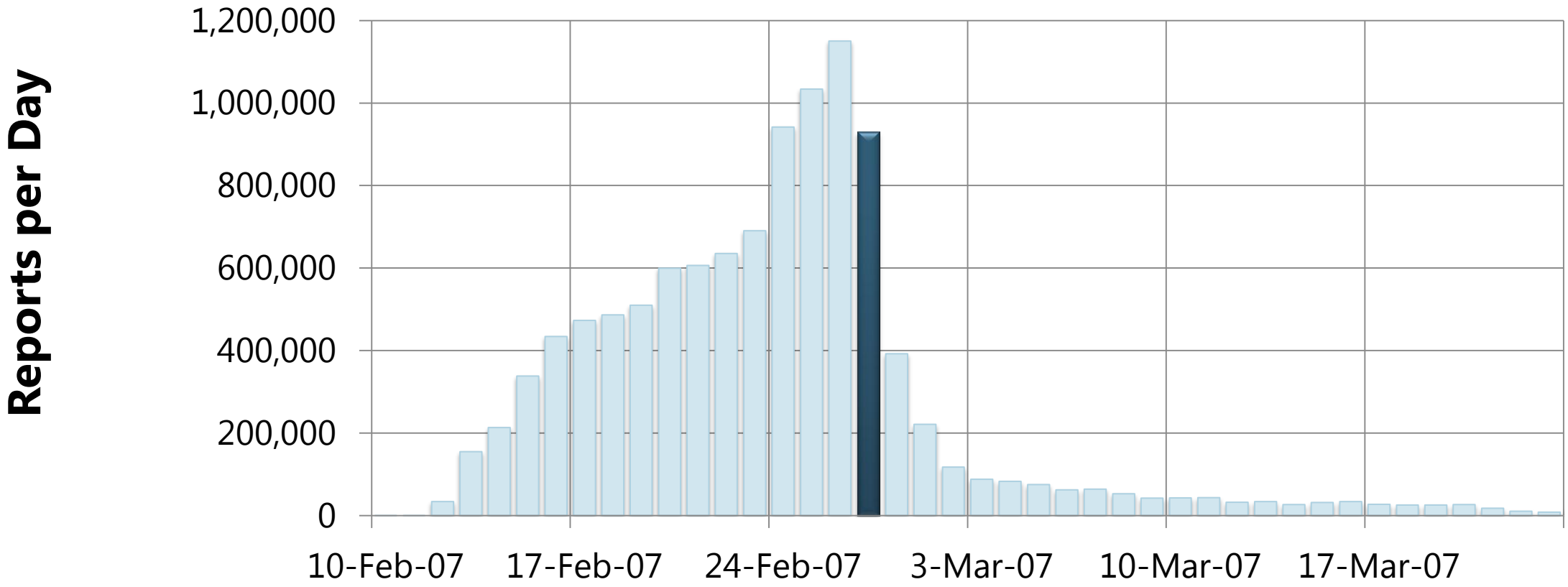
Hardware: Processor Bug



WER helped fix hardware error

Manufacturer **could** have caught this earlier with WER

Renos Malware



Early detection w/o user action (renos, blaster, slammer, etc.)

WER **scales** to handle global events

3. SOFTWARE IN THE BROWSER

Types of software

1.Only a browser, e.g., Explorer, Firefox, Safari

2. A platform and language(s), e.g., a CLI implementation and C# or the JVM

3. An integrated development environment (IDE), e.g., Visual Studio or Eclipse

Browser based software

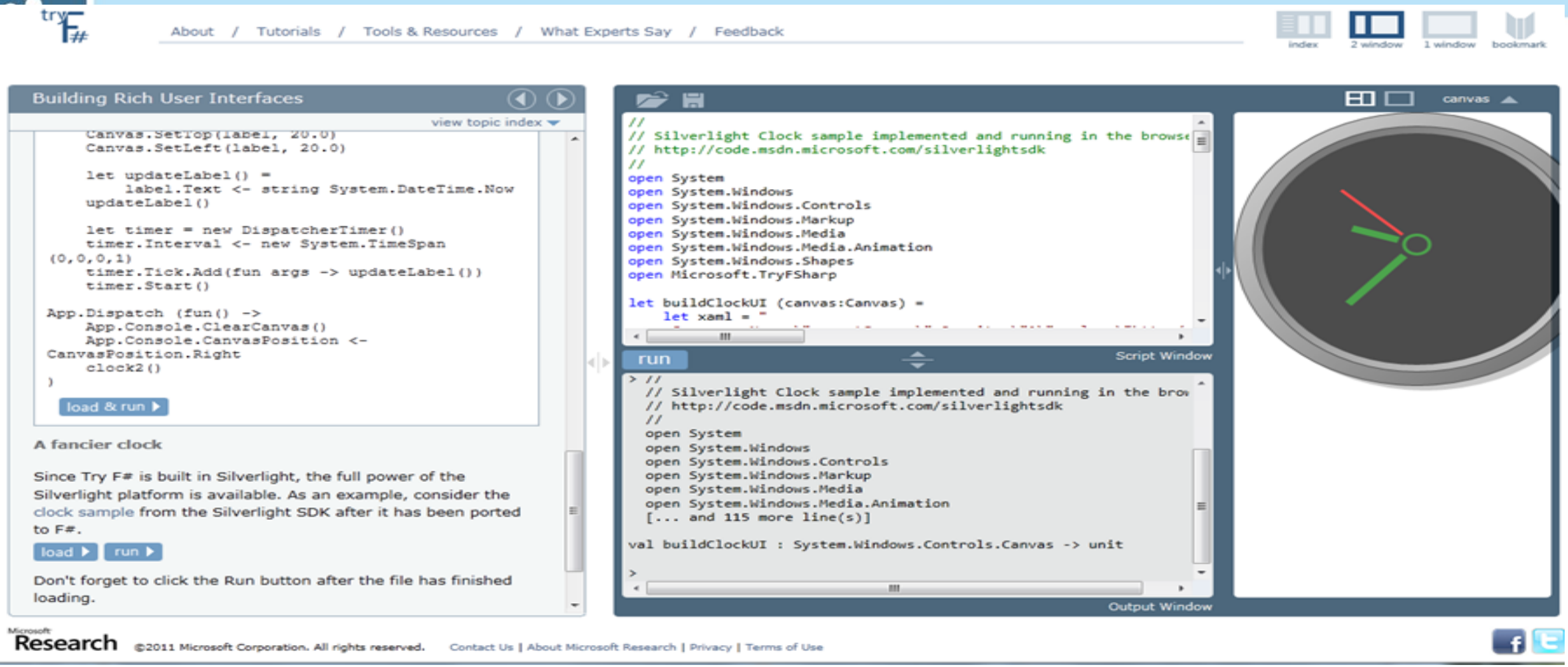
Sandbox Approach

- Download a Silverlight/ Moonlight control with a complete compiler
- All interaction directed to the control from the browser
- Computation on client
- Con: Effort to create the system
- Pro: No additional hardware needed

Server Approach

- Maintain a server (or cloud) presence
- All interaction directed to the server from the browser
- Computation on server
- Con: Scalability issue
- Pro: Can gather data on usability

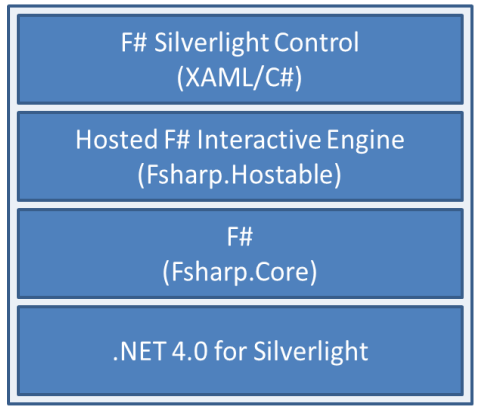
www.tryfsharp.org



Don Syme, Dean Guo, Christophe Poulain, Joe Pamer, Laurent le Brun, Nigel Horspool, Judith Bishop, 2010

Inside the site

Try F# Control
Built in Silverlight



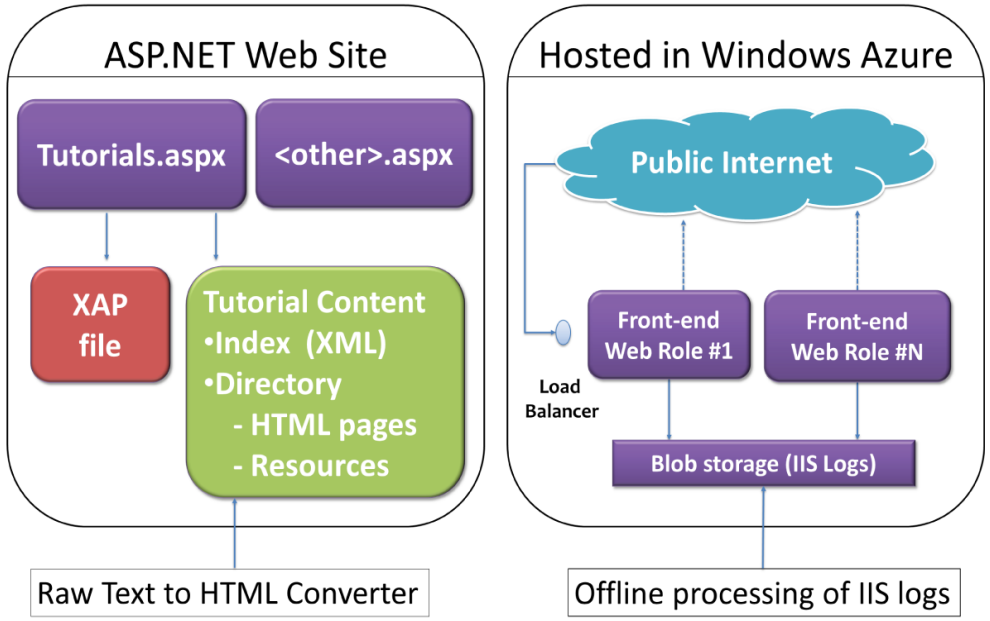
Compiled into XAP file

- Manifest
- DLL for our control
- FSharp.Hostable.dll
- FSharp.Core.dll
- .NET Silverlight DLLs

Control is scriptable:

```
string Script { get; set }
string Output { get; }
void ClearOutput()
int LinesToEcho { get; set; }
Canvas Canvas { get; }
void ClearCanvas()
CanvasPosition CanvasPosition { get; set; }
void LoadFromString(string script)
void LoadFromUrl(string url)
void LoadFromFile()
bool IsDirty { get; set }
void Save()
void Execute(string script)
void Cancel()
void Reset()
```

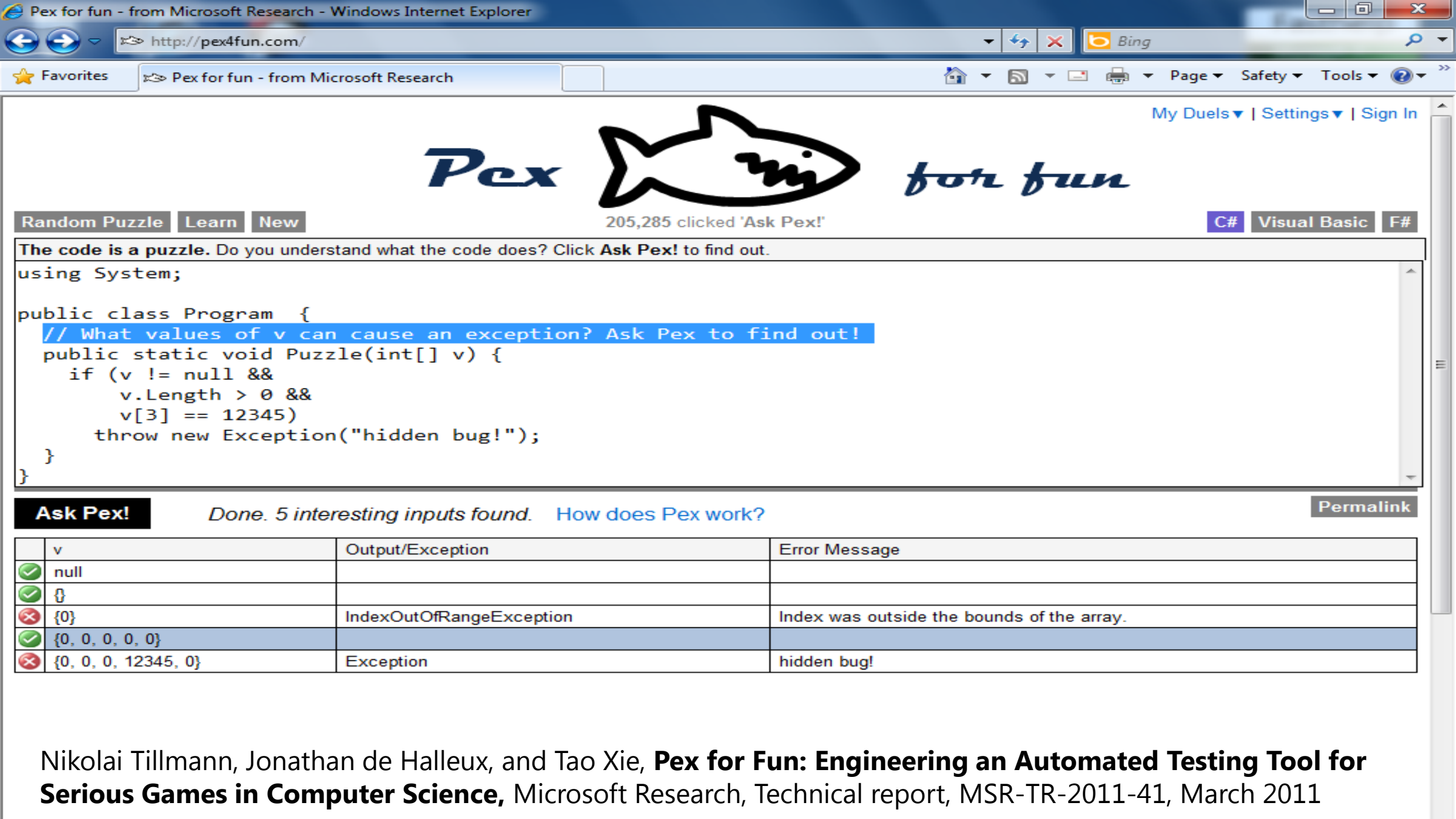
Try F# Web Site
Deployed in Windows Azure



Demo of F#

www.Pex4fun.com

- Pex - Visual Studio 2010 Power Tool developed by Microsoft Research to help unit testing of .NET applications.
  - Can be launched from the command line and run as Type 2 or Type 3 software.
- Pex4Fun
  - radically simplified version of the fully featured Pex
  - accessed via a browser
  - all the work happens on one of Microsoft Research servers
  - creates a game out of unit testing by providing existing or user entered code puzzles in C#, Visual Basic, or F#
  - Users determine from the unit tests what code needs to be added or changed.



Nikolai Tillmann, Jonathan de Halleux, and Tao Xie, **Pex for Fun: Engineering an Automated Testing Tool for Serious Games in Computer Science,** Microsoft Research, Technical report, MSR-TR-2011-41, March 2011

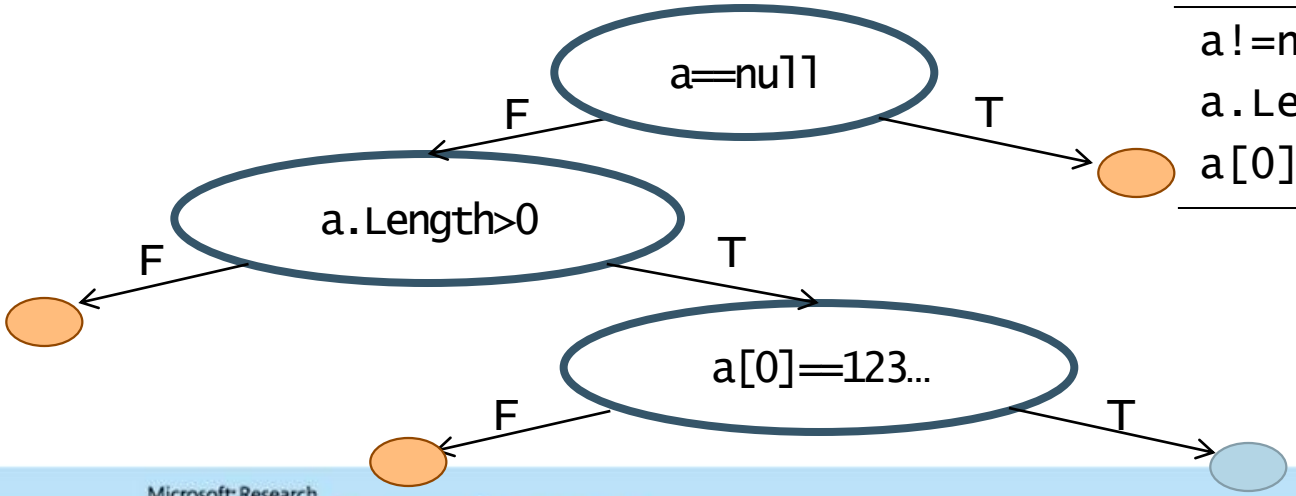
Dynamic Symbolic Execution in Pex

- Generates test data systematically

```
void CoverMe(int[] a)
{
  if (a == null) return;
  if (a.Length > 0)
    if (a[0] == 1234567890)
      throw new Exception("bug");
}
```

Constraints to solve	Input	Observed constraints
	null	a==null
a!=null	{}	a!=null && !(a.Length>0)
a!=null && a.Length>0	{0}	a==null && a.Length>0 && a[0]!=1234567890
a!=null && a.Length>0 && a[0]==123456890	{123..}	a==null && a.Length>0 && a[0]==1234567890

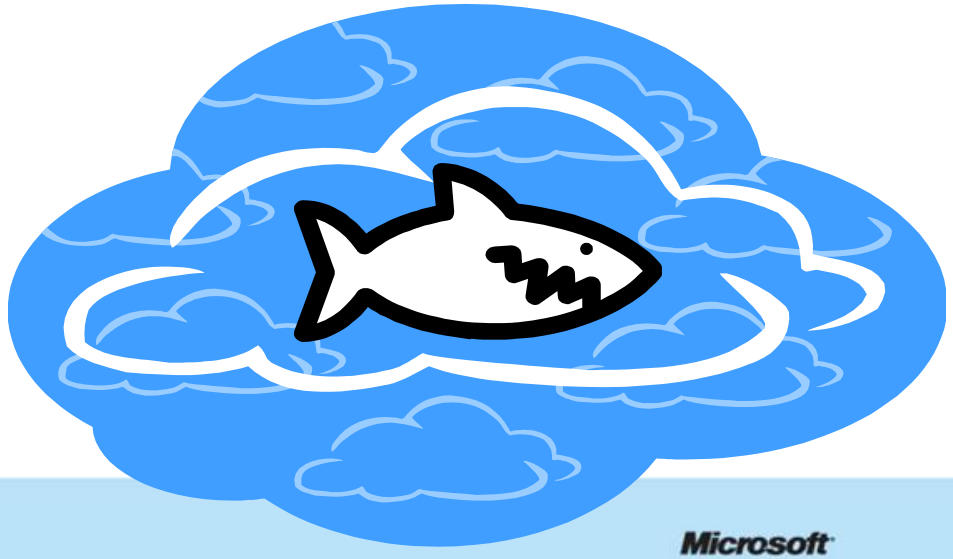
Done: There is no path left.



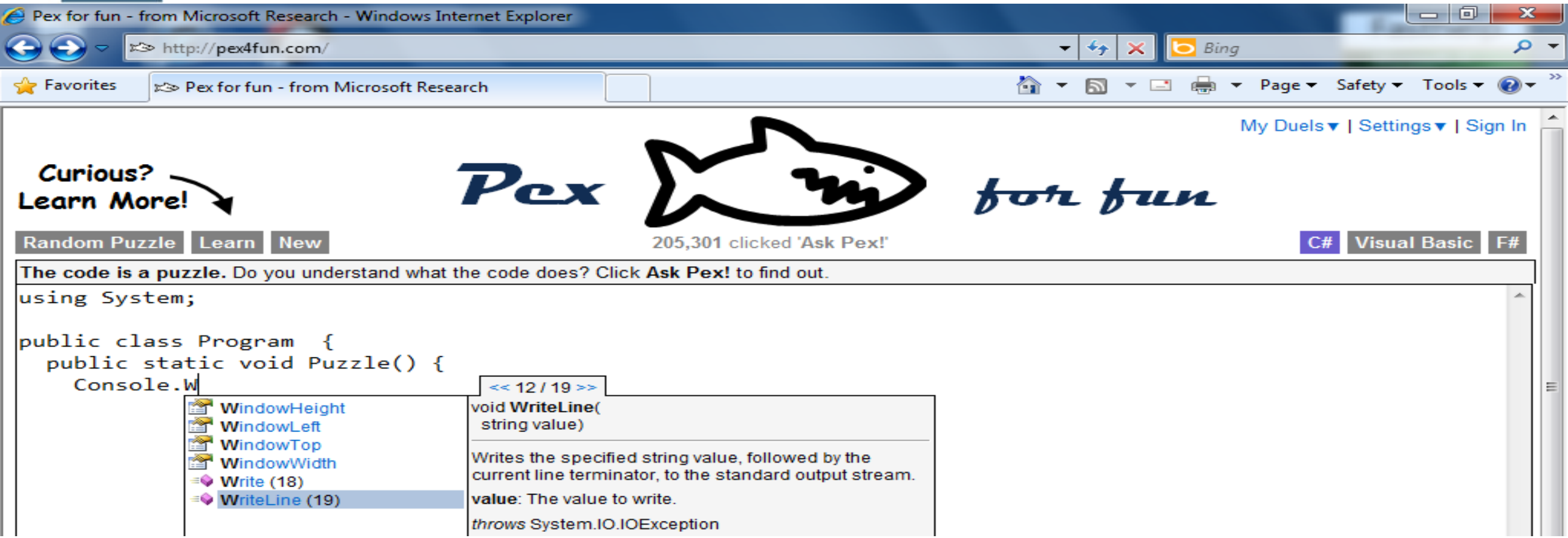
http://pex4fun.com/HowDoesPexWork

Dynamic Symbolic Execution Summary

- "Ask Pex" sends code to cloud
- Code is compiled and analyzed in cloud
- Dynamic Symbolic Execution automatically finds relevant interesting test inputs that achieve high code coverage
- Results are shown in browser



Auto-Completion



Rise4Fun.com

Research in Software Engineering (RiSE)

RiSE coordinates Microsoft's Research in Software Engineering in Redmond, USA. Our mission is to advance the state of the art in Software Engineering and Programming Languages, and to bring those advances to Microsoft's business.

http://rise4fun.com, our tools in **your** browser.

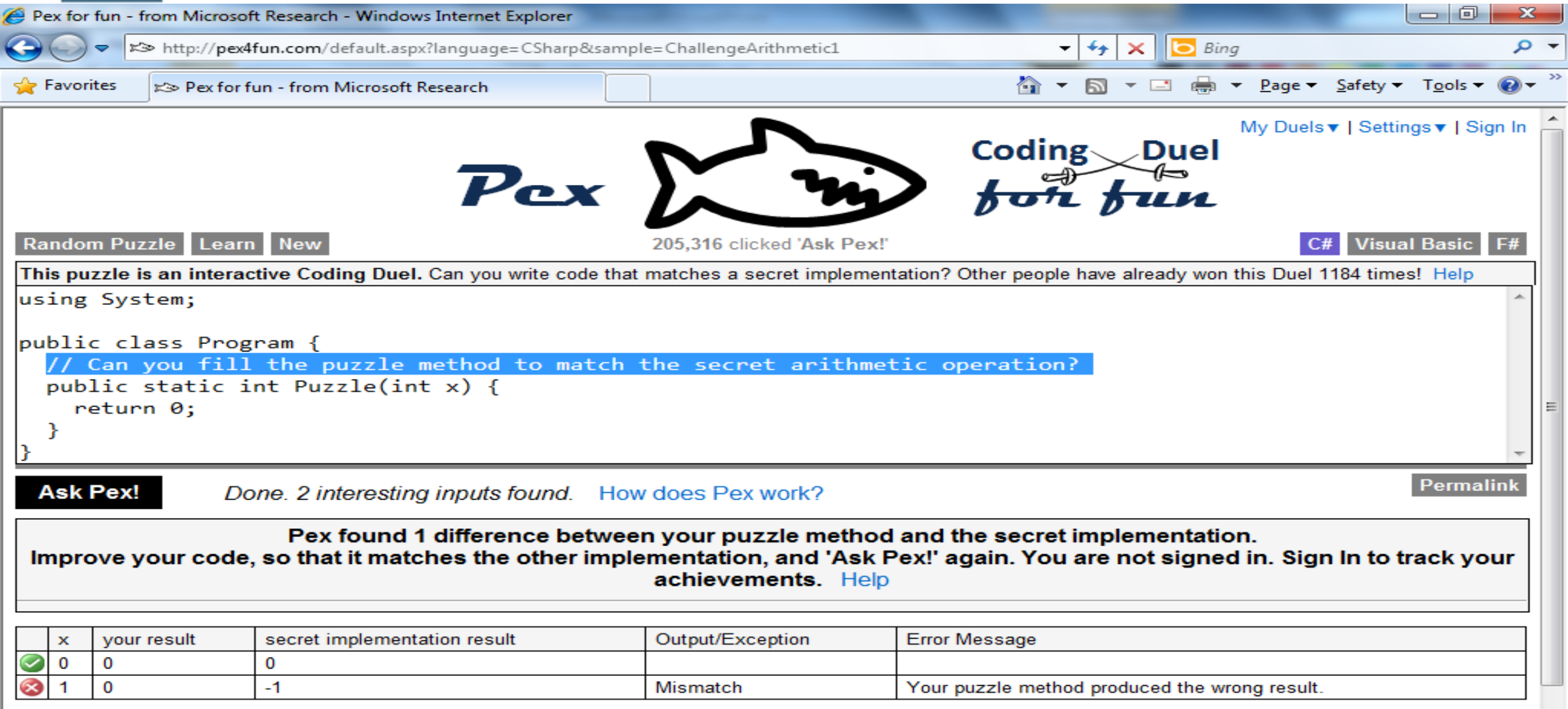
see all...

Using video clips on Channel9

4. SOCIAL EXPERIENCE

Try F# on Facebook

Coding Duels

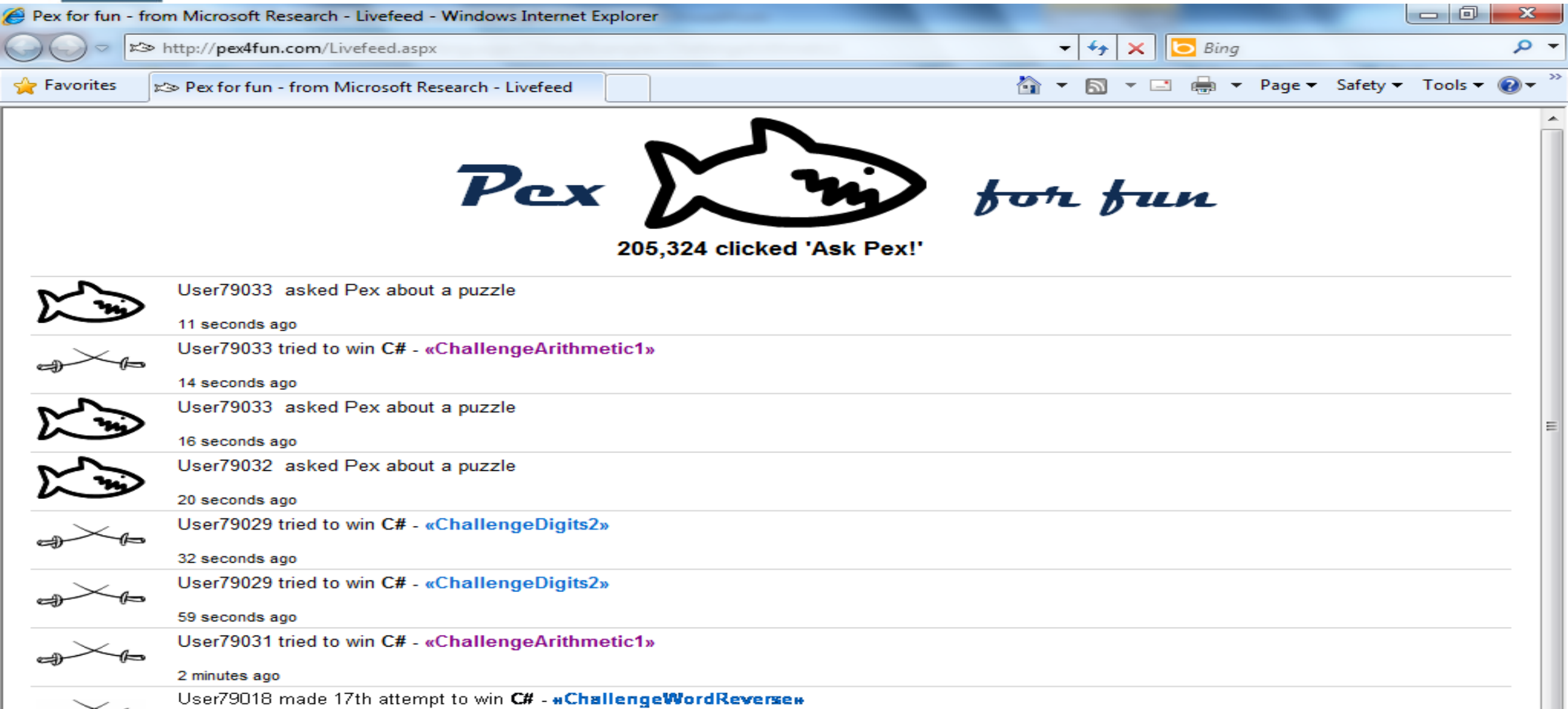


	x	your result	secret implementation result	Output/Exception	Error Message
✓	0	0	0		
✗	1	0	-1	Mismatch	Your puzzle method produced the wrong result.

Coding Duels - Fun and Engaging

- Iterative gameplay
- Adaptive
- Personalized
- No cheating
- Clear winning criterion

Social Experience



Social Experience Summary

- Community
- High score lists, leaderboard
- Live feed

http://pex4fun.com/Community.aspx
http://pex4fun.com/Livefeed.aspx

Demo of Pex4Fun

5. SOCIAL ASPECTS IN THE LARGE

MS Products' (Bing) History is Huge

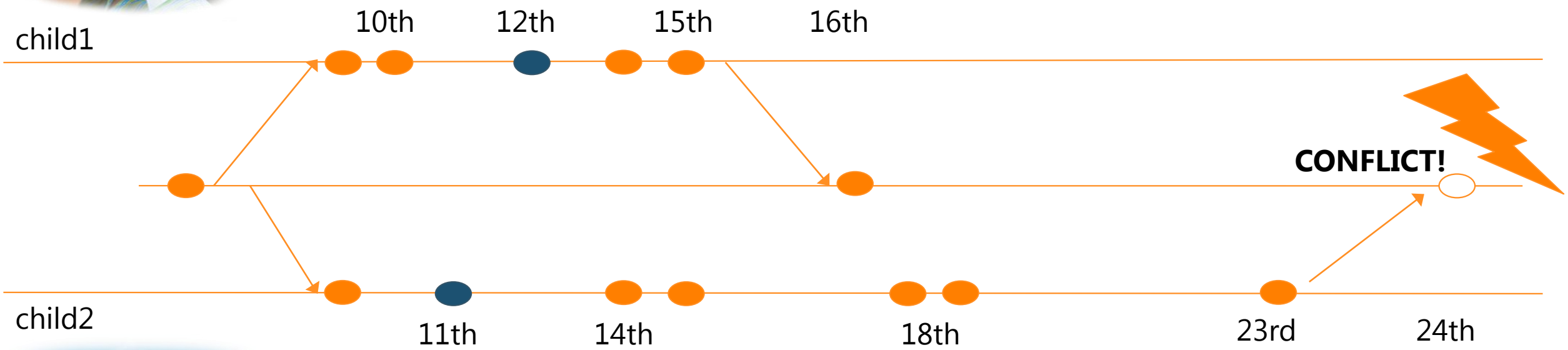
- Over one million change-sets.
- One change-set can contain 200K check-ins.
  - e.g., branch creations
- +700 branches, 6M different files, 150M different depot paths, 450M different revisions, 150M check-ins, 130M integrations
- Almost impossible to crawl incrementally
  - At least not in 3 months!

Speculative Analysis

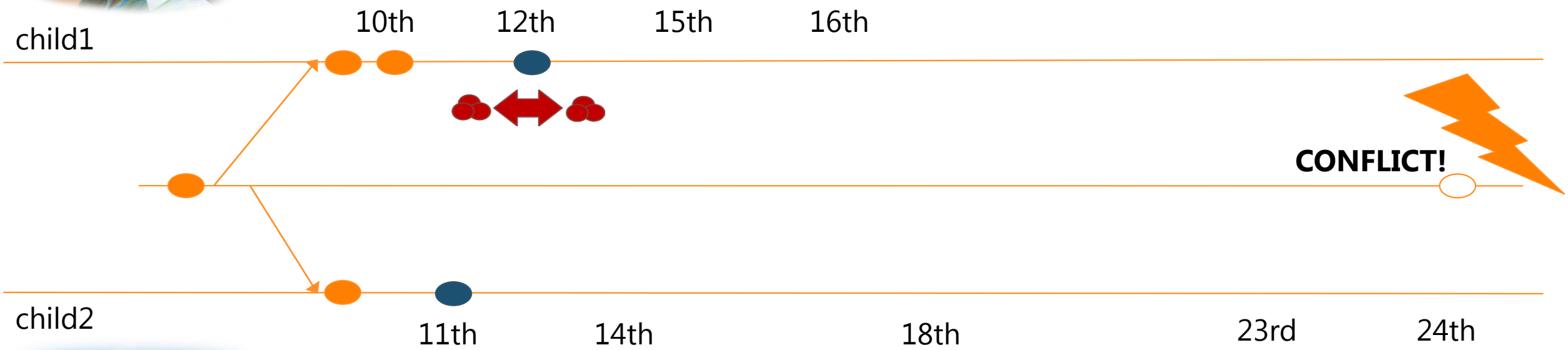
- *Predict* what the user might do
- *Do* that action in advance, store the results
- If the user *chooses* to do it,
- *Present* the pre-calculated information

- Aim: increase developer awareness with precise information

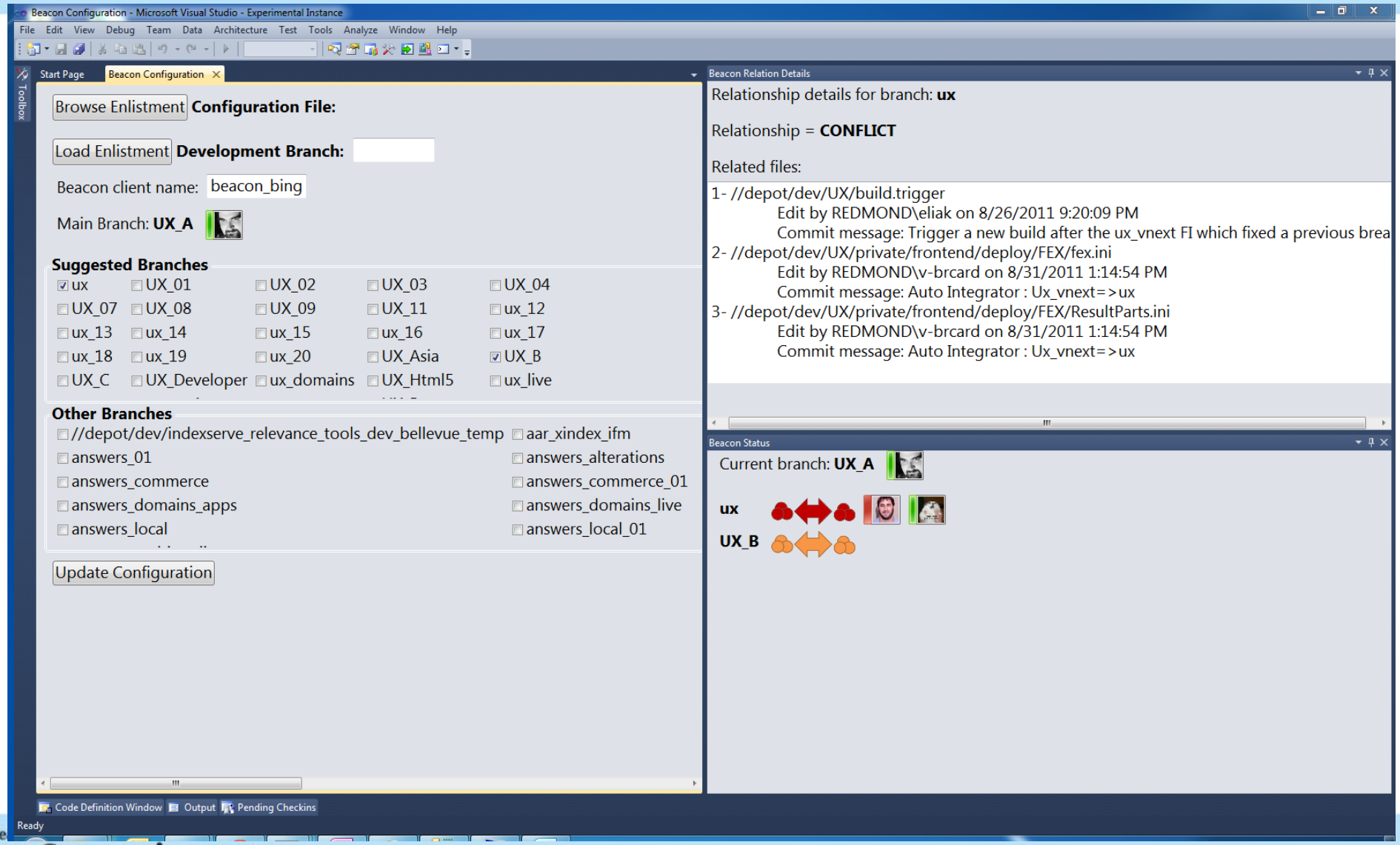
Development History (Recall)



Development History (with BEACON)



BEACON Example



6. SOFTWARE DEVELOPMENT KITS

Software Development Kits

- Used for access to
  - Proprietary hardware and their drivers
  - Large proprietary data

Research.microsoft.com/cs

Project Hawaii	Kinect SDK	Web-NGram
• On WP7 • Services execute in the cloud (Azure) • OCR, Speech to text etc • **WP7 phones loaned to universities worldwide** • C#	• Drivers and rich APIs for raw sensor streams and human motion tracking • **Kinect unit is $150** • Available to academics • C++, C#, VB	• Content and model types • N-gram availability to 5 • Training size: ***All*** **documents indexed by Bing in the en-us market** • Updated Periodically

Arjmand Samuel, Stewart Tansley, Evelyne Viegas

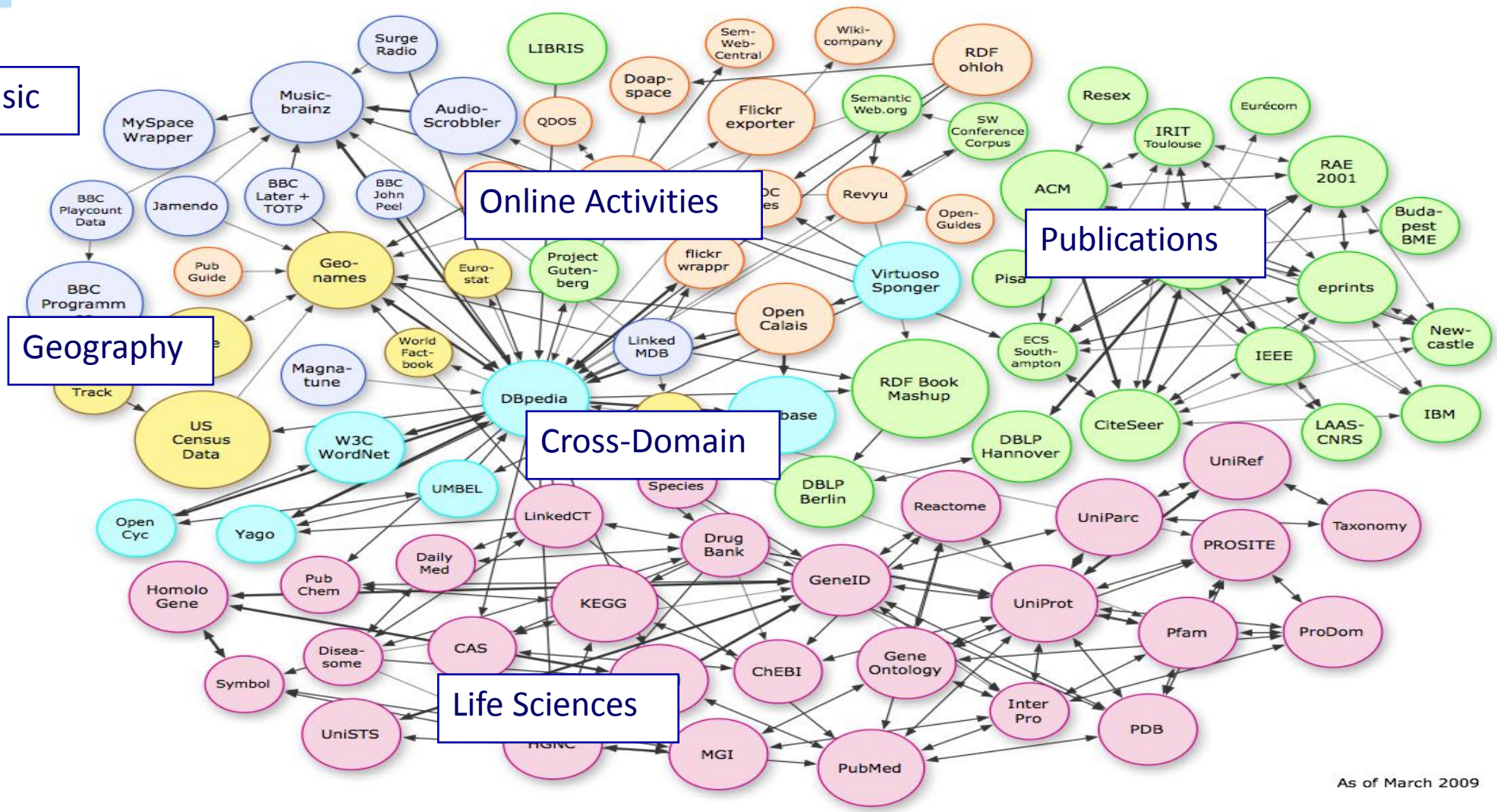
Web N-Gram in Public Beta

Single Tag Cloud

Multi Tag Cloud

Ref: Dr. Li Ding, Rensselaer Polytechnic Institute

Linked Data



Attribution: Richard Cyganiak

Concept Search

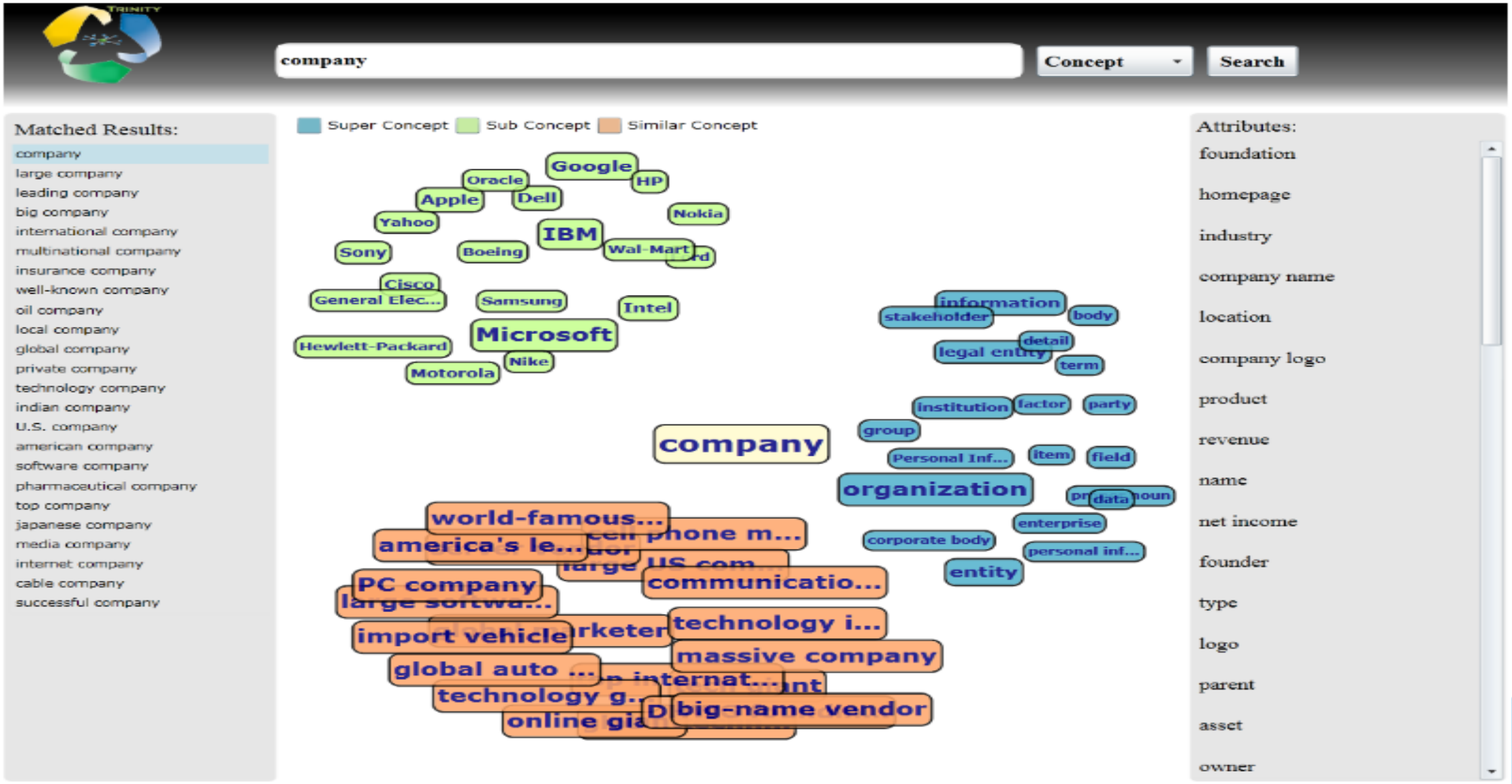
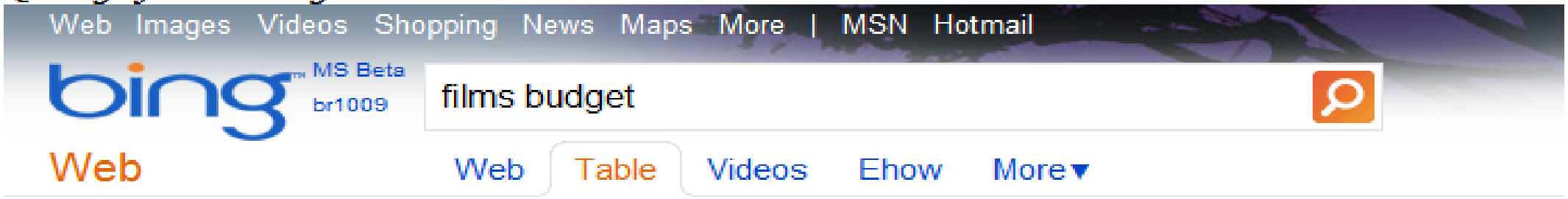


Table understanding

Query: films budget



Shrink

Shrink table

Year	Movie	Worldwide gross	Budget	Distributor	Director
1977	Star Wars	$ 782400000	$ 11000000	20th Century Fox	George Lucas
1997	Titanic	$ 1848813795	$ 200000000	Paramount Pictures	James Cameron
1993	Jurassic Park	$ 914691118	$ 95000000	Universal Studios	Steven Spielberg
1995	Toy Story	$ 365000000	$ 90000000	Walt Disney Pictures	John Lasseter
1972	The Godfather	$ 245066411	$ 6000000	Paramount Pictures	Francis Ford Coppola
2009	Avatar	$ 2606954237	$ 237000000	20th Century Fox	James Cameron
1975	Jaws	$ 470600000	$ 7000000	Universal Studios	Steven Spielberg
1996	Independence Day	$ 816969268	$ 75000000	20th Century Fox	Roland Emmerich
1998	Armageddon	$ 553709788	$ 140000000	Touchstone Pictures	Michael Bay

SEIF Awards

research.microsoft.com/seif

Software Engineering Innovation Foundation (SEIF) with MSR's-RiSE group supports academic research in software engineering

- Technologies, tools and practices
- Teaching methods

Awards of US$ 25,000 for

- seed-funding larger initiatives,
- proofs of concept, or
- demonstrations of feasibility

In association with the RiSE and Mobile groups at Microsoft Research

- **LATAM Awards**
  - 2010 – 3 out of 12
  - 2011 – 1 out of 10
  - In open global competition
- **Connected to major conferences a workshop and an event**

SEIF DAY, July 18, 2012

Conclusion - best practices

- Software engineering is analytic and deals with huge data
- Technology transfer to academia by moving to
  - Browsers
  - SDKs
  - New devices
- Encouraging community through social media