THE WEB
CHANGES EVERYTHING

Jaime Teevan, Microsoft Research
Project Tuva’s Feynman Lectures
Gates’ Gift to Lifelong Learning
Welcome to Microsoft Research! This Web Space is designed to inform you about our group. If you are looking for something (or someone) in particular, please select the Search button at the top or bottom of this page for quick access to information contained in this space. To learn more about MSR, check out the Microsoft Research Introduction page.

To help structure this space we have divided our research areas into three primary "Research Themes" and listed the various research groups according to theme. In addition to individual group pages, each theme has an overview page with thumbnail descriptions of each group's activities. Independent researchers are listed on the Colleagues page.
William Henry "Bill" Gates III (born October 28, 1955) is an American business magnate, philanthropist, and chairman of Microsoft, the software company he founded with Paul Allen. He is ranked consistently one of the world's wealthiest people and the wealthiest overall as of 2009. During his career at Microsoft, Gates held the positions of CEO and chief software architect, and remains the largest individual shareholder with more than 8 percent of the common stock. He has also authored or co-authored several books.

Gates is one of the best-known entrepreneurs of the personal computer revolution. Although he is admired by many, a number of industry insiders criticize his business tactics, which they consider anti-competitive, an opinion which has in some cases been upheld by the courts (see Criticism of Microsoft). In the later stages of his career, Gates has pursued a number of philanthropic endeavors, donating large amounts of money to various charitable organizations and scientific research programs through the Bill & Melinda Gates Foundation, established in 2000.

Bill Gates stepped down as chief executive officer...
Bill Gates

For other persons named Bill Gates, see Bill Gates (disambiguation).

William Henry Gates III (born October 28, 1955) [2] is an American business magnate, philanthropist, and chairman[3] of Microsoft, the software company he founded with Paul Allen. During his career at Microsoft, Gates held the positions of CEO and chief software architect, and remains the largest individual shareholder with more than 9 percent of the common stock.[8] He has also authored or co-authored several books.

Gates is one of the best-known entrepreneurs of the personal computer revolution. Although he is admired by many, a large number of industry insiders criticize his business tactics, which they consider anti-competitive, an opinion which has in some cases been upheld by the courts.[9][7] In the later stages of his career, Gates has pursued a number of philanthropic endeavors, donating large amounts of money to various charitable organizations and scientific research programs through the Bill & Melinda Gates Foundation, established in 2000.

Bill Gates stepped down as chief executive officer of Microsoft in January, 2000. He remained as chairman and created the position of chief software architect. In June, 2006, Gates
William Henry "Bill" Gates III (born October 28, 1955) is an American business magnate, philanthropist, and chairman of Microsoft, the software company he founded with Paul Allen. He is ranked consistently one of the world's wealthiest people and the wealthiest overall as of 2009. During his career at Microsoft, Gates held the positions of CEO and chief software architect, and remains the largest individual shareholder with more than $8 percent of the common stock. He has also authored or co-authored several books.

Gates is one of the best-known entrepreneurs of the personal computer revolution. Although he is admired by many, a large number of industry insiders criticize his business tactics, which they consider anti-competitive, an opinion which has in some cases been upheld by the courts (see Criticism of Microsoft). In the later stages of his career, Gates has pursued a number of philanthropic endeavors, donating large amounts of money to various charitable organizations and scientific research programs through the Bill & Melinda Gates Foundation, established in 2000. Bill Gates stepped down as chief executive officer of Microsoft in 2000 and then as chairman in 2014, but remains actively involved as a global citizen.
William Henry "Bill" Gates III (born October 28, 1955) is an American business magnate, philanthropist, and chairman of Microsoft, the software company he founded with Paul Allen. He is ranked consistently one of the world's wealthiest people and the wealthiest overall as of 2009. During his career at Microsoft, Gates held the positions of CEO and chief software architect, and is the largest individual shareholder with 1.7 percent of the common stock. He has co-authored several books.

The best-known entrepreneurs of the computer revolution. Although he is often associated with large companies, his business tactics, which they claim are competitive, an opinion which has been upheld by the courts (see Intel vs. Microsoft). In the later stages of his career, he has pursued a number of philanthropic endeavors, donating large amounts of money to charitable organizations and programs through the Bill & Melinda Gates Foundation, established in 2000.
The Web Changes Everything

<table>
<thead>
<tr>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
</table>

Content Changes
Bill Gates

Bill Gates
From Wikipedia, the free encyclopedia

William Henry "Bill" Gates III (born October 28, 1955) is an American business magnate, philanthropist, and chairman of Microsoft, the software company he founded with Paul Allen. He is ranked consistently one of the world's wealthiest people and the wealthiest overall as of 2009. During his career at Microsoft, Gates held the positions of CEO and chief software architect, and remains the largest individual shareholder with more than 9% percent of the common stock. He has also authored or co-authored several books.

Gates is one of the best-known entrepreneurs of the personal computer revolution. Although he is admired by many, a large number of industry insiders criticize his business tactics, which they consider anti-competitive, an opinion which has in some cases been upheld by the courts (see Criticism of Microsoft). In the later stages of his career, Gates has pursued a number of philanthropic endeavors, donating large amounts of money to various charitable organizations and scientific research programs through the Bill & Melinda Gates Foundation, established in 2000.

Bill Gates stepped down as chief executive officer
The Web Changes Everything

Content Changes

January  February  March  April  May  June  July  August  September

People Revisit

Today’s tools focus on the present

But there’s so much more information available!
The Web Changes Everything

Content Changes

- Large scale Web crawl over time
  - Revisited pages
    - 55,000 pages crawled hourly for 18+ months
  - Judged pages (relevance to a query)
    - 6 million pages crawled every two days for 6 months
Measuring Web Page Change

- **Summary metrics**
  - Number of changes
  - Time between changes
  - Amount of change

Top level pages change by more and faster than pages with long URLs. .edu and .gov pages do not change by very much or very often. News pages change quickly, but not as drastically as other types of pages.
Measuring Web Page Change

- **Summary metrics**
  - Number of changes
  - Time between changes
  - Amount of change

- **Change curves**
  - Fixed starting point
  - Measure similarity over different time intervals

![Diagram showing change curves with a knot point at the Dice Similarity vs Time from starting point graph.](image)
Measuring Within-Page Change

- DOM structure changes
- Term use changes
  - Divergence from norm
    - cookbooks
    - frightfully
    - merrymaking
    - ingredient
    - latkes
  - Staying power in page
Accounting for Web Dynamics

- Avoid problems caused by change
  - Caching, archiving, crawling
- Use change to our advantage
  - Ranking
    - Match term’s staying power to query intent
  - Snippet generation

---

**Tom Bosley - Wikipedia, the free encyclopedia**

Bosley died at 4:00 a.m. of heart failure on October 19, 2010, at a hospital near his home in Palm Springs, California. … His agent, Sheryl Abrams, said Bosley had been battling lung cancer.

[en.wikipedia.org/wiki/tom_bosley](en.wikipedia.org/wiki/tom_bosley)
Revisititation on the Web

Content Changes

People Revisit

What’s the last Web page you visited?
Revisitation on the Web

- **Revisitation patterns**
  - Log analysis
    - Browser logs for revisitation
    - Query logs for re-finding
  - User survey for intent

Content Changes

People Revisit

What’s the last Web page you visited?
Measuring Revisitation

- Summary metrics
  - Unique visitors
  - Visits/user
  - Time between visits

- Revisitation curves
  - Revisit interval histogram
  - Normalized
Four Revisitation Patterns

- **Fast**
  - Hub-and-spoke
  - Navigation within site
- **Hybrid**
  - High quality fast pages
- **Medium**
  - Popular homepages
  - Mail and Web applications
- **Slow**
  - Entry pages, bank pages
  - Accessed via search engine
Search and Revisitation

- Repeat query (33%)
  - web science conference

- Repeat click (39%)
  - http://websci11.org
  - Query → websci 11

- Lots of repeats (43%)
  - Many navigational
How Revisitation and Change Relate

Content Changes

January February March April May June July August September

People Revisit

Why did you revisit the last Web page you did?
How Revisitation and Change Relate

Content Changes

People Revisit

Why did you revisit the last Web page you did?
Possible Relationships

- Interested in change
  - Monitor

- Effect change
  - Transact

- Change unimportant
  - Find

- Change can interfere
  - Re-find
Understanding the Relationship

- Compare summary metrics
- Revisits: Unique visitors, visits/user, interval
- Change: Number, interval, similarity

<table>
<thead>
<tr>
<th></th>
<th>Number of changes</th>
<th>Time between changes</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 visits/user</td>
<td>172.91</td>
<td>133.26</td>
<td>0.82</td>
</tr>
<tr>
<td>3 visits/user</td>
<td>200.51</td>
<td>119.24</td>
<td>0.82</td>
</tr>
<tr>
<td>4 visits/user</td>
<td>234.32</td>
<td>109.59</td>
<td>0.81</td>
</tr>
<tr>
<td>5 or 6 visits/user</td>
<td>269.63</td>
<td>94.54</td>
<td>0.82</td>
</tr>
<tr>
<td>7+ visits/user</td>
<td>341.43</td>
<td>81.80</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Comparing Change and Revisit Curves

- Three pages
  - New York Times
  - Woot.com
  - Costco

- Similar change patterns

- Different revisitation
  - NYT: Fast
  - Woot: Medium
  - Costco: Slow

![Graph showing change and revisitation curves for NYT, Woot, and Costco](image)
Within-Page Relationship

- Page elements change at different rates
- Pages revisited at different rates
- Resonance can serve as a filter for interesting content
Coalition in Pakistan Moves to Impeach Musharraf
By JANE PERLEZ 12:19 PM ET
Pakistan’s usually fractious coalition government moved decisively for the first time on Thursday to impeach President Pervez Musharraf.
• Times Topics: Pervez Musharraf | Pakistan
Post a Comment | Read (33)
E.P.A. Won’t Ease Requirements for Ethanol in Gas
By MATTHEW L. WALD 23 minutes ago
The agency denied a request to cut the quota for the use of ethanol in cars, concluding that the goal of reducing the nation’s reliance on oil trumps any effect on food prices.
In Anthrax Scientist’s E-Mail, Hints of Delusions
By ERIC LIPTON
The e-mail messages of Bruce E. Ivins provide a narrative of his paranoid episodes, dating to 2003.
• The Lede: Reactions to the Case 12:22 PM ET
• F.B.I. Presents Anthrax Case, Saying Scientist Acted Alone 10 minutes ago
• Times Topics: Anthrax
SPORTS OF THE TIMES
Raising the Bar and Raising Elbows
By GEORGE VECSEY 11:07 AM ET
Increased physicality in women’s basketball will not give a black eye to the entire sport. However, it most certainly gives a black eye to Penny Taylor of Australia on Tuesday.
INTERACTIVE SCHEDULE
MORE OLYMPICS NEWS
• Men’s Soccer: U.S. 1, Japan 0
• Bush Arrives in Beijing
• Silver Keeps Her Place on U.S. Relay Team
• Federer Arrives, Soon to Lose His No. 1 Status
• Rings Blog
• Complete Coverage »
Judge Orders Detroit Mayor Jailed
By NICK BUNKLEY 39 minutes ago
Sometimes The Past Is Hard To Let Go

"...and when she told me, I just couldn't deal. So I left, just like I always did. Just like my father."

Tom, thank you for sharing all that. Have one of these tissues. Do you feel like you've made a forward step?

"I do, Jack. I really do. Can I maybe have a second tissue?"

As many as you like, Tom, you don't even need to ask. Let's hear from someone else now. HP Compaq 2GHz Celeron 17" Widescreen Notebook, tell us more about your loss.

Well, with me, it was the teasing. See, I'm in the tech industry, which is a really high stress field. And, you know, growing up, everyone in a position of authority says, you know, play the hand you're dealt. But I've got a 2.0GHz Intel Celeron 550 Processor with 1MB L2 cache and 533MHz front side bus, and most people think that's nothing. But, you know, what are they doing? Playing Bejeweled on their lunch break and writing out memos in work? Maybe looking for um, pee oh are on?
Building Support for Web Dynamics

Content Changes

January  February  March  April  May  June  July  August  September

People Revisit
Exposing Change with Diff-IE

Interesting Features of Diff-IE

- Always on
- In-situ
- New to you
- Non-intrusive

Examples of Diff-IE in Action

Expected New Content

Monitor


Twitter

What are you doing?

Latest: Coding data and thinking about how people phrase the questions they ask about 1 month ago

Home

brynn Study by Rutgers Prof (@informer) Reveal Two Types of Twitter Users, "Informers" and "Miformers": http://bit.ly/PUCvD (via @edchi)
1 minute ago from web

emax and i sped up the ticker to not melt my poor desktop, now we show everyone's surfing activity in real time ~ http://bit.ly/eyebrowseticker
26 minutes ago from web

emax if you have an eyebrowse account, log in + checkout your homepage ~ @zamiang made it beautiful and full of dots! http://bit.ly/eyebrows
29 minutes ago from web

dtunkelanq My blog post today at #thenoisychannel
Unexpected Important Content

Serendipitous Encounters

Organized by Type * Subject * Date * Internal

**ETEREED JOURNAL ARTICLES:**


**ETEREED CONFERENCE PAPERS:**


Unexpected Unimportant Content

Understand Page Dynamics

### Attend to Activity

**'13: Eastside Mothers & More Expo.** See Chapter News and Events for more information. **Yeah, baby!!**

![Image](http://bit.ly/DiffIE)

---

#### Others & More Message Board | Discussion Forums | Sequencing

| 1] 2 3 ... 9 | Mark Topics as Read for this Board | Notify of replies | Start new topic | Post

<table>
<thead>
<tr>
<th>Subject</th>
<th>Started by</th>
<th>Replies</th>
<th>Views</th>
<th>Last post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Business « 12 »</td>
<td>Della Dawn</td>
<td>21</td>
<td>206</td>
<td>February 17, 2009, 05 PM by Lori Poythress</td>
</tr>
<tr>
<td>Defining Sequencing</td>
<td>Miti Morris</td>
<td>0</td>
<td>93</td>
<td>November 10, 2009, 06 PM by Miti Morris</td>
</tr>
<tr>
<td>Back to work Part Time! new</td>
<td>Sue Monahan</td>
<td>8</td>
<td>43</td>
<td>Today at 03:36:34 PM by Lynn Woods</td>
</tr>
<tr>
<td>Aldersgate Christian Preschool - HIRING new</td>
<td>Heather Ross</td>
<td>0</td>
<td>17</td>
<td>Today at 03:01:42 PM by Heather Ross</td>
</tr>
<tr>
<td>Do you know anyone in Italy or Sweden?</td>
<td>Virginia Menting</td>
<td>3</td>
<td>70</td>
<td>Today at 07:00:21 AM by Virginia Menting</td>
</tr>
<tr>
<td>Got an interview and I'm freaking out... « 1 2 ... 6 7 »</td>
<td>Trina Sooy</td>
<td>94</td>
<td>962</td>
<td>June 08, 2009, 06:54 PM by Kyla Carlson</td>
</tr>
<tr>
<td>Academia or Industry? Finding the Right Fit</td>
<td>Jaime Teevan</td>
<td>7</td>
<td>105</td>
<td>May 26, 2009, 09:26:12 AM by Trina Sooy</td>
</tr>
<tr>
<td>Any members work at Amazon.com?</td>
<td>Kathryn Pizzo</td>
<td>10</td>
<td>188</td>
<td>May 08, 2009, 08:20:21 PM by JJ Galley</td>
</tr>
<tr>
<td>Time Magazine on SAHMs getting back into workforce</td>
<td>Kathryn Pizzo</td>
<td>2</td>
<td>75</td>
<td>April 17, 2009, 07:34:18 AM by Erin Proctor</td>
</tr>
<tr>
<td>Freelance Resource</td>
<td>Kathryn Pizzo</td>
<td>0</td>
<td>41</td>
<td>April 01, 2009, 06:54:59 PM by Kathryn Pizzo</td>
</tr>
<tr>
<td>Graphic designer or Admin?</td>
<td>Heather Ross</td>
<td>0</td>
<td>45</td>
<td>March 20, 2009, 07:53:30 PM by Heather Ross</td>
</tr>
</tbody>
</table>
Installing DiffIE

- Installation location: \asibuild\Drops\CLUES\latest\DiffIE\DiffIE.msi
  Note: Save this MSI to your machine and stop IE before installing DiffIE.
- Announcement alias: If you download DiffIE, please join the “DiffIE Announcements” distribution list so that we can keep you informed about new versions and other important changes.
- Feedback alias: diffiefb@microsoft.com

Using DiffIE

Once installed, you should see the DiffIE toolbar in IE (shown below on Vista):
Expected

- New Content

Monitor

Unexpected

- Important Content

- Edit

- Page Dynamics

- Unexpected

- Unimportant Content

Attend to Activity

Serendipitous Encounter

Monitor
Find Expected New Content

Studying **Diff-IE**

Content Changes

[SURVEY](http://bit.ly/DiffIE)
- How often do pages change?  
- How often do you revisit?

People Revisit

---

Install **Diff-IE**
Seeing Change Changes Web Use

- Changes to perception
  - *Diff-IE* users become more likely to notice change
  - Provide better estimates of how often content changes

- Changes to behavior
  - *Diff-IE* users start to revisit more
    - Revisited pages more likely to have changed
    - Changes viewed are bigger changes
  - Content gains value when history is exposed

The Web Changes Everything

Content Changes

Web content changes provide valuable insight

Relating revisitation and change enables us to

- Identify pages for which change is important
- Identify interesting components within a page

People revisit and re-find Web content

Explicit support for Web dynamics can impact how people use and understand the Web
Thank you.

**Web Content Change**

**Web Page Revisitation**

**Relating Change and Revisitation**

**Studying Diff-IE**
August 4, 2006: Logs released to academic community
- 3 months, 650 thousand users, 20 million queries
- Logs contain anonymized User IDs

August 7, 2006: AOL pulled the files, but already mirrored

August 9, 2006: New York Times identified Thelma Arnold
- “A Face Is Exposed for AOL Searcher No. 44177749”
- Queries for businesses, services in Lilburn, GA (pop. 11k)
- Queries for Jarrett Arnold (and others of the Arnold clan)
- NYT contacted all 14 people in Lilburn with Arnold surname
- When contacted, Thelma Arnold acknowledged her queries

August 21, 2006: 2 AOL employees fired, CTO resigned

September, 2006: Class action lawsuit filed against AOL
Example: AOL Search Dataset

- **August 4, 2006**: Logs released to academic community
  - 3 months, 650 thousand users, 20 million queries
  - Logs contain anonymized User IDs
- **August 7, 2006**: AOL pulled the files, but already mirrored
- **August 9, 2006**: New York Times identified Thelma Arnold
  - “A Face Is Exposed for AOL Searcher No. 4417749”
  - Queries for businesses, services in Lilburn, GA (pop. 11k)
  - Queries for Jarrett Arnold (and others of the Arnold clan)
  - NYT contacted all 14 people in Lilburn with Arnold surname
  - When contacted, Thelma Arnold acknowledged her queries
- **August 21, 2006**: 2 AOL employees fired, CTO resigned
- **September, 2006**: Class action lawsuit filed against AOL

<table>
<thead>
<tr>
<th>AnonID</th>
<th>Query</th>
<th>QueryTime</th>
<th>ItemRank</th>
<th>ClickURL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234567</td>
<td>jitp</td>
<td>2006-04-04 18:18:18</td>
<td>1</td>
<td><a href="http://www.jitp.net/">http://www.jitp.net/</a></td>
</tr>
<tr>
<td>1234567</td>
<td>computational social science</td>
<td>2006-04-24 09:19:32</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1234567</td>
<td>computational social science</td>
<td>2006-04-24 09:20:04</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1234567</td>
<td>seattle restaurants</td>
<td>2006-04-24 09:25:50</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1234567</td>
<td>perlman montreal</td>
<td>2006-04-24 10:15:14</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

- When contacted, Thelma Arnold acknowledged her queries
- **August 21, 2006**: 2 AOL employees fired, CTO resigned
- **September, 2006**: Class action lawsuit filed against AOL

*AnonID, Query, QueryTime, ItemRank, ClickURL*
Example: AOL Search Dataset

- August 4, 2006: Logs released to academic community
  - 3 months, 650 thousand users, 20 million queries
  - Logs contain anonymized User IDs
- August 7, 2006: AOL pulled the files, but already mirrored
- August 9, 2006: New York Times identified Thelma Arnold
  - “A Face Is Exposed for AOL Searcher No. 4417749”
  - Queries for businesses, services in Lilburn, GA (pop. 11k)
  - Queries for Jarrett Arnold (and others of the Arnold clan)
  - NYT contacted all 14 people in Lilburn with Arnold surname
  - When contacted, Thelma Arnold acknowledged her queries
- August 21, 2006: 2 AOL employees fired, CTO resigned
- September, 2006: Class action lawsuit filed against AOL

<table>
<thead>
<tr>
<th>AnonID</th>
<th>Query</th>
<th>QueryTime</th>
<th>ItemRank</th>
<th>ClickURL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234567</td>
<td>jitp</td>
<td>2006-04-04 18:18:18</td>
<td>1</td>
<td><a href="http://www.jitp.net/">http://www.jitp.net/</a></td>
</tr>
<tr>
<td>1234567</td>
<td>computational social scinece</td>
<td>2006-04-24 09:19:32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1234567</td>
<td>computational social science</td>
<td>2006-04-24 09:20:04</td>
<td>2</td>
<td><a href="http://socialcomplexity.gmu.edu/phd.php">http://socialcomplexity.gmu.edu/phd.php</a></td>
</tr>
<tr>
<td>1234567</td>
<td>seattle restaurants</td>
<td>2006-04-24 09:25:50</td>
<td>2</td>
<td><a href="http://seattletimes.nwsource.com/rets">http://seattletimes.nwsource.com/rets</a></td>
</tr>
</tbody>
</table>

...
Example: AOL Search Dataset

- Other well known AOL users
  - User 927 *how to kill your wife*
  - User 711391 *i love alaska*
    - [http://www.minimovies.org/documentaires/view/ilovealaska](http://www.minimovies.org/documentaires/view/ilovealaska)

- Anonymous IDs do not make logs anonymous
  - Contain directly identifiable information
    - Names, phone numbers, credit cards, social security numbers
  - Contain indirectly identifiable information
    - Example: Thelma’s queries
    - Birthdate, gender, zip code identifies 87% of Americans
Example: Netflix Challenge

- October 2, 2006: Netflix announces contest
  - Predict people’s ratings for a $1 million dollar prize
  - 100 million ratings, 480k users, 17k movies
  - Very careful with anonymity post-AOL

- May 18, 2008: Data de-anonymized
  - Paper published by Narayanan & Shmatikov
  - Uses background knowledge from IMDB
  - Robust to perturbations in data

- December 17, 2009: Doe v. Netflix

- March 12, 2010: Netflix cancels second competition
Example: Netflix Challenge

- October 2, 2006: Netflix announces contest
  - Predict people’s ratings for a $1 million prize
  - 100 million ratings, 480k users, 17k movies
  - Very careful with anonymity post-AOL
- May 18, 2008: Data de-anonymized
  - Paper published by Narayanan & Shmatikov
  - Uses background knowledge from IMBD
  - Robust to perturbations in data
- December 17, 2009: Doe v. Netflix
- March 12, 2010: Netflix cancels second competition
Example: Netflix Challenge

- **October 2, 2006**: Netflix announces contest
  - Predict people’s ratings for a $1 million prize
  - 100 million ratings, 480k users, 17k movies
  - Very careful with anonymity post-AOL

- **May 18, 2008**: Data de-anonymized
  - Narayanan & Shmatikov
  - Uses background knowledge from IMDB
  - Robust to perturbations in data

- **December 17, 2009**: Doe v. Netflix
  - March 12, 2010: Netflix cancels second competition
Example: Netflix Challenge

- October 2, 2006: Netflix announces contest
  - Predict people’s ratings for a $1 million prize
  - 100 million ratings, 480k users, 17k movies
  - Very careful with anonymity post-AOL

- May 18, 2008: Data de-anonymized
  - Paper published by Narayanan & Shmatikov
  - Uses background knowledge from IMDB
  - Robust to perturbations in data

- December 17, 2009: Doe v. Netflix

- March 12, 2010: Netflix cancels second competition

Ratings
1: [Movie 1 of 17770]
12, 3, 2006-04-18 [UserID, Rating, Date]
1234, 5, 2003-07-08 [UserID, Rating, Date]
2468, 1, 2005-11-12 [UserID, Rating, Date]
...

Movie Titles
...
10120, 1982, “Bladerunner”
17690, 2007, “The Queen”
...

All customer identifying information has been removed; all that remains are ratings and dates. This follows our privacy policy. . .
Even if, for example, you knew all your own ratings and their dates you probably couldn’t identify them reliably in the data because only a small sample was included (less than one tenth of our complete dataset) and that data was subject to perturbation.