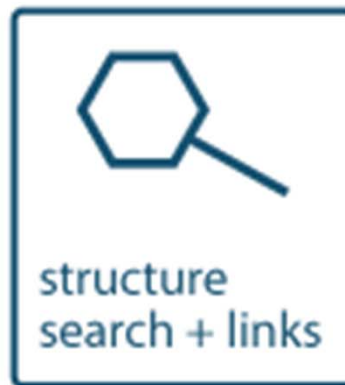
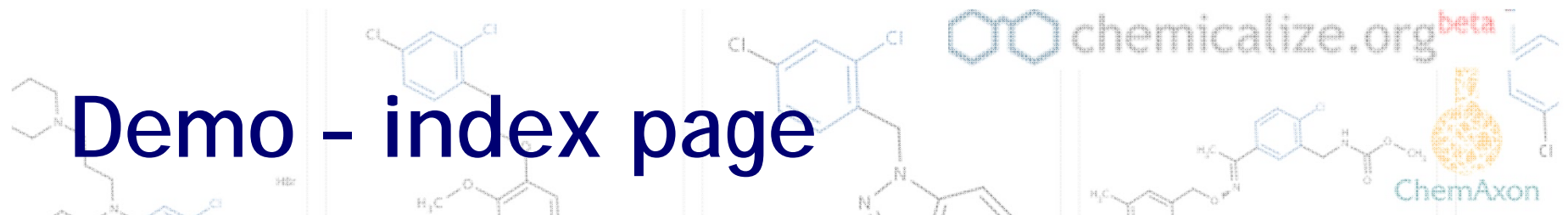


# Adding structures to Web pages and data to structures

Alex Allardyce  
ChemAxon





# Demo - index page

- Lay out input box
- Recently chemicalized, recent queries...
- Drag and drop structure images
- Help, about

Type a chemical name or URL to begin

draw upload

- chemicalize
- Calculate »
- Search »
- Web Search »

Example: <http://www.chemicalize.org/>

# Demo - chemicalizing a Web page

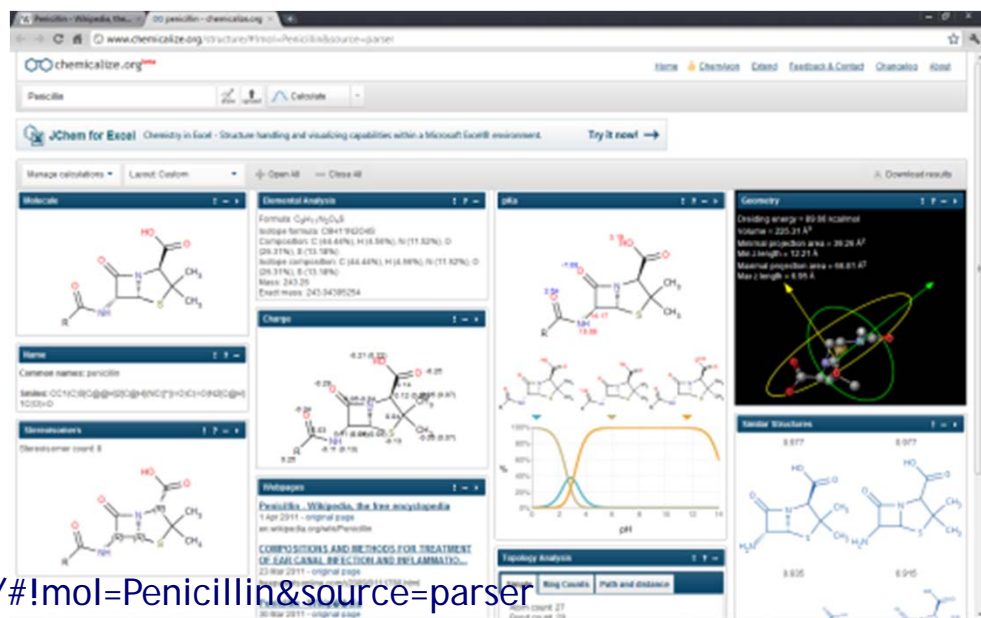
- URL paste
  - Structure images
  - TOC and links
  - Properties link from mouse over image
  - Download
  - Links work

The screenshot shows the chemicalize.org interface. At the top, there is a navigation bar with the logo and the text "chemicalize.org beta". Below this, a row of chemical structures is displayed, each with a number (1-10) and a small icon. The main content area shows a Wikipedia article for "Penicillin". The article text is partially visible, mentioning "Penicillin (sometimes abbreviated PCN or PENICILIN)" and "Penicillin antibiotics are historically significant because they are the first drugs that were effective...". A table of contents is visible on the right side of the article, listing sections such as "1 Mechanism of action", "2 Variants in clinical use", "3 Structure", "4 Biosynthesis", "5 Production", "6 Pharmacology", "7 History", "7.1 Discovery", "7.2 Medical application", "7.3 Mass production", and "7.4 Developments from penicillin". A chemical structure of Penicillin is shown on the right side of the article, with a "Report error" link below it.

Example: <http://www.chemicalize.org/?url=http%3A%2F%2Fen.wikipedia.org%2Fwiki%2FPenicillin>

# Demo - Structure based predictions

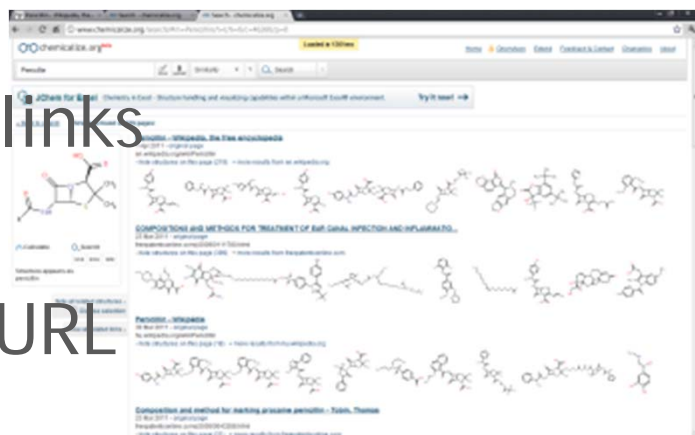
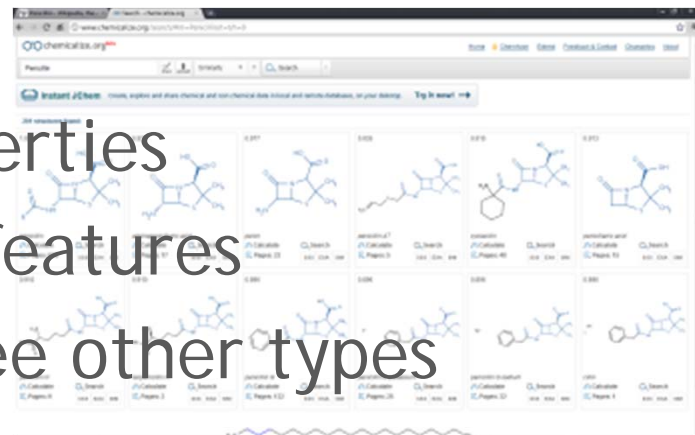
- Properties
  - Manage views, move boxes
  - Open MarvinView from double click on any structure image
  - Calculate on demand
  - Download results



Example: <http://www.chemicalize.org/structure/#!mol=Penicillin&source=parser>

# Demo - Structure search

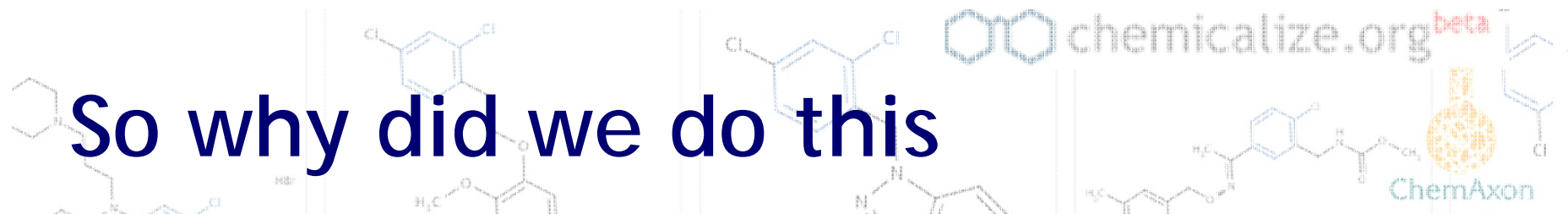
- Chem search pages
  - Search from Calculate properties
  - Open Marvin, power query features
  - Similarity default search, see other types
- Choose a structure
  - List of URL's, chemicalized links
  - Show structures
  - Combine chem search with URL
  - Download results



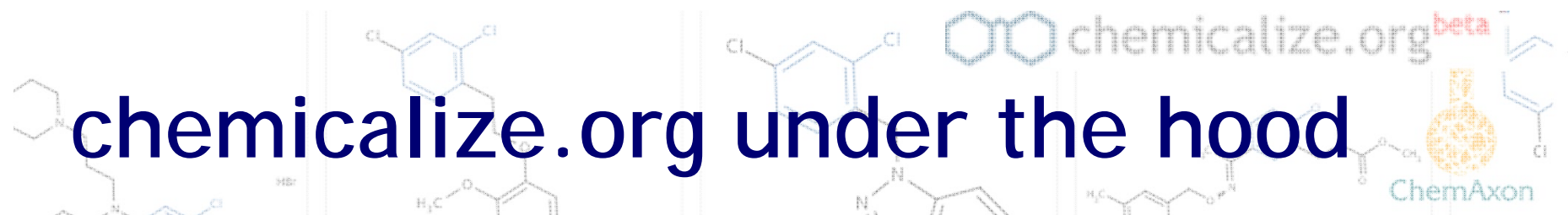
Examples: 1. <http://www.chemicalize.org/search/#m=Penicillin/t=t/h=0>

2. <http://www.chemicalize.org/search/#m=Penicillin/t=t/h=0/c=46260/p=0>





- There is a lot of content on the web
- Useful + increase visibility/utility of chemical structures
- Creates user interest in this type of functionality and so demand for chemistry and content for publishers
- Lets us develop directly with end users:
  - Functionality/feature development
  - GUI usability
  - Crowd sourced bug fixing “Report Error” for naming.
- Pushing state of the art
  - Browser tech (svg, chunking, reducing calls)
  - ChemAxon tech (on the web, must be superfast, finalise features)
- We love cheminformatics “cheminfomaniacs”



# chemicalize.org under the hood

web application (15kloc):

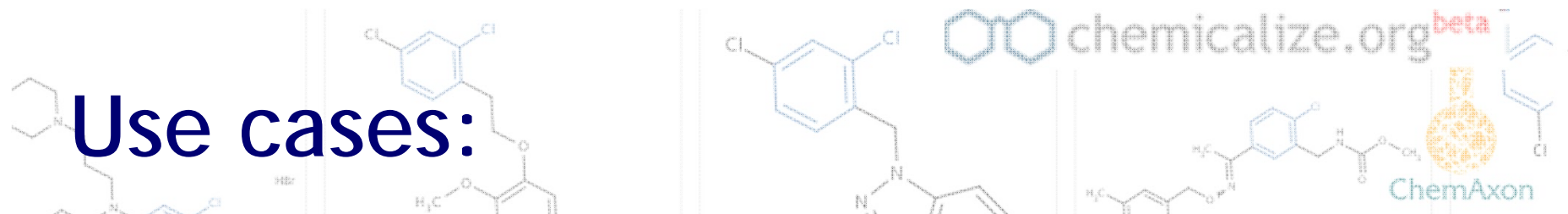
- MySQL: DB engine - structure/text storage
- ChemAxon bits: see below
- Apache Tomcat - servlet container with code logic
- jQuery + Plugins - UI interactions with code logic
  - A fair bit of home grown (46% of code) here



The background features several chemical structures and logos. On the left, there's a blue chemical structure. In the center, a benzene ring with two chlorine atoms and a side chain. To the right, the 'chemicalize.org beta' logo with a blue hexagonal icon. Further right, another chemical structure with a benzene ring and a side chain. On the far right, the 'ChemAxon' logo with a yellow flask icon.

# ChemAxon bits

- [Marvin](#): structure editor, viewer, image generation
- [Name <> structure, Document to Structure](#): parsing, dictionaries and lexing IUPAC names
- [JChem Base](#), [JChem Web Services](#), [Standardizer](#), [MCES](#): structure database, duplicate checking, structure search, web services layer, canonicalization, hit highlighting
- [Calculator Plugins](#): structure based predictions like pKa, logP, logD, charge, HBDA, tautomer, stereoisomers, etc.  
Notable combined predictions yield argument results - like "Lipinski-likeness" etc



- Wanted to know the logP of...
- What are the structures for known drugs ([http://en.wikipedia.org/wiki/List\\_of\\_drugs](http://en.wikipedia.org/wiki/List_of_drugs))
- Seeing structures in relation to the name
- All wikipedia pages with a “chembox” have been indexed by chemicalize.org so can be searched by structure search (sub structure, similar, exact)
- See all similar structures (and names) for any similar structure : sildenafil = viagra, tadalafil, vardenafil, udenafil ...
- Draw a structure and see it's name
- Automatically chemicalize my blog (WordPress plugin)



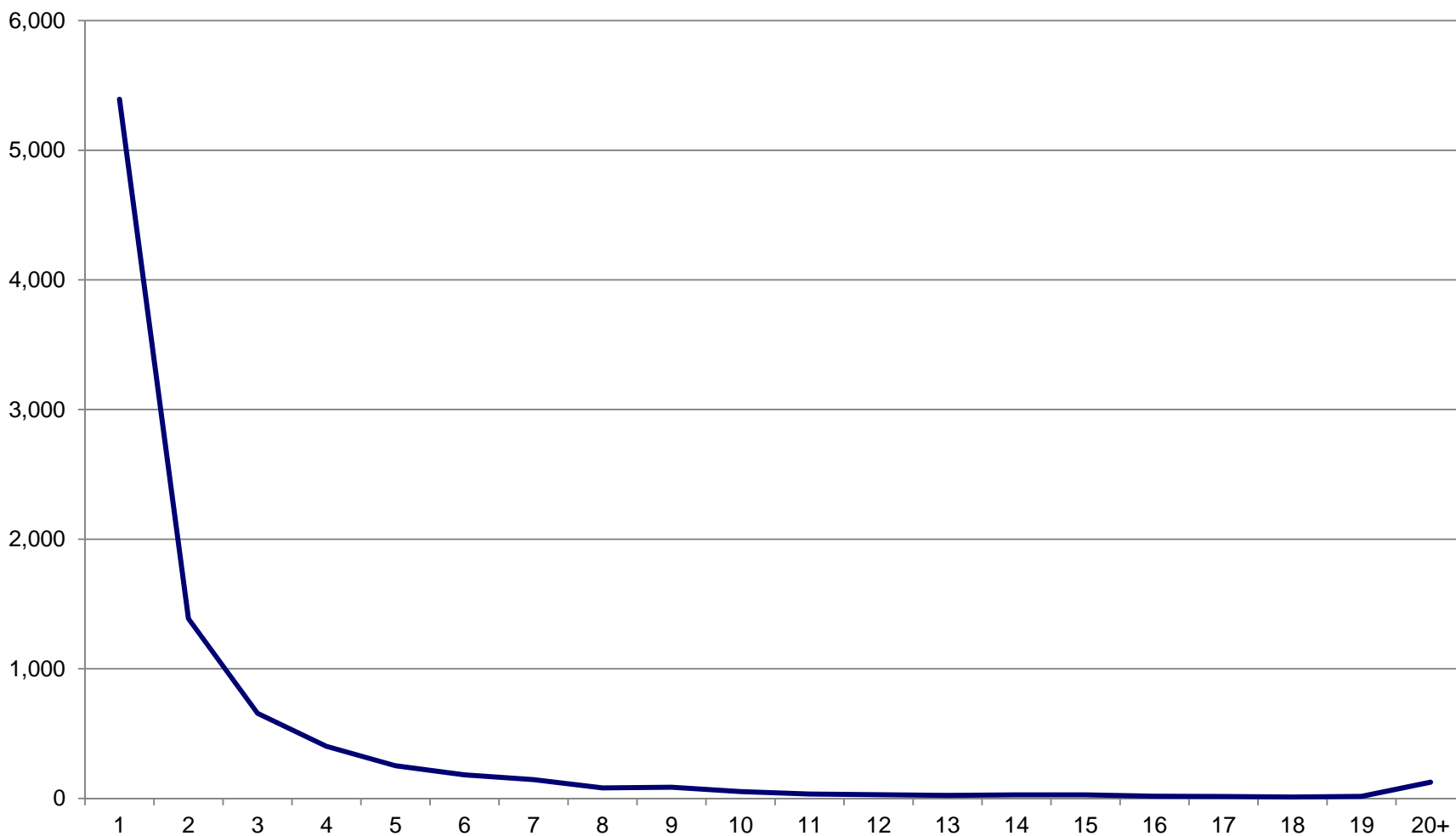
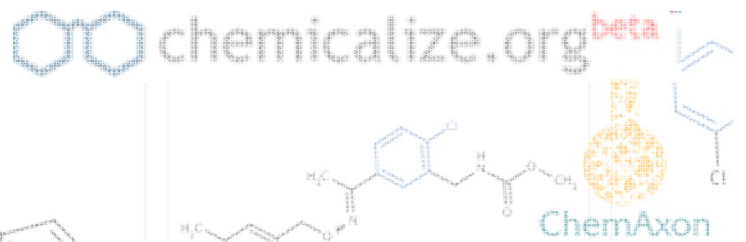
# Stats: Raw numbers

(Apr 1, 2010 - Mar 25, 2011)

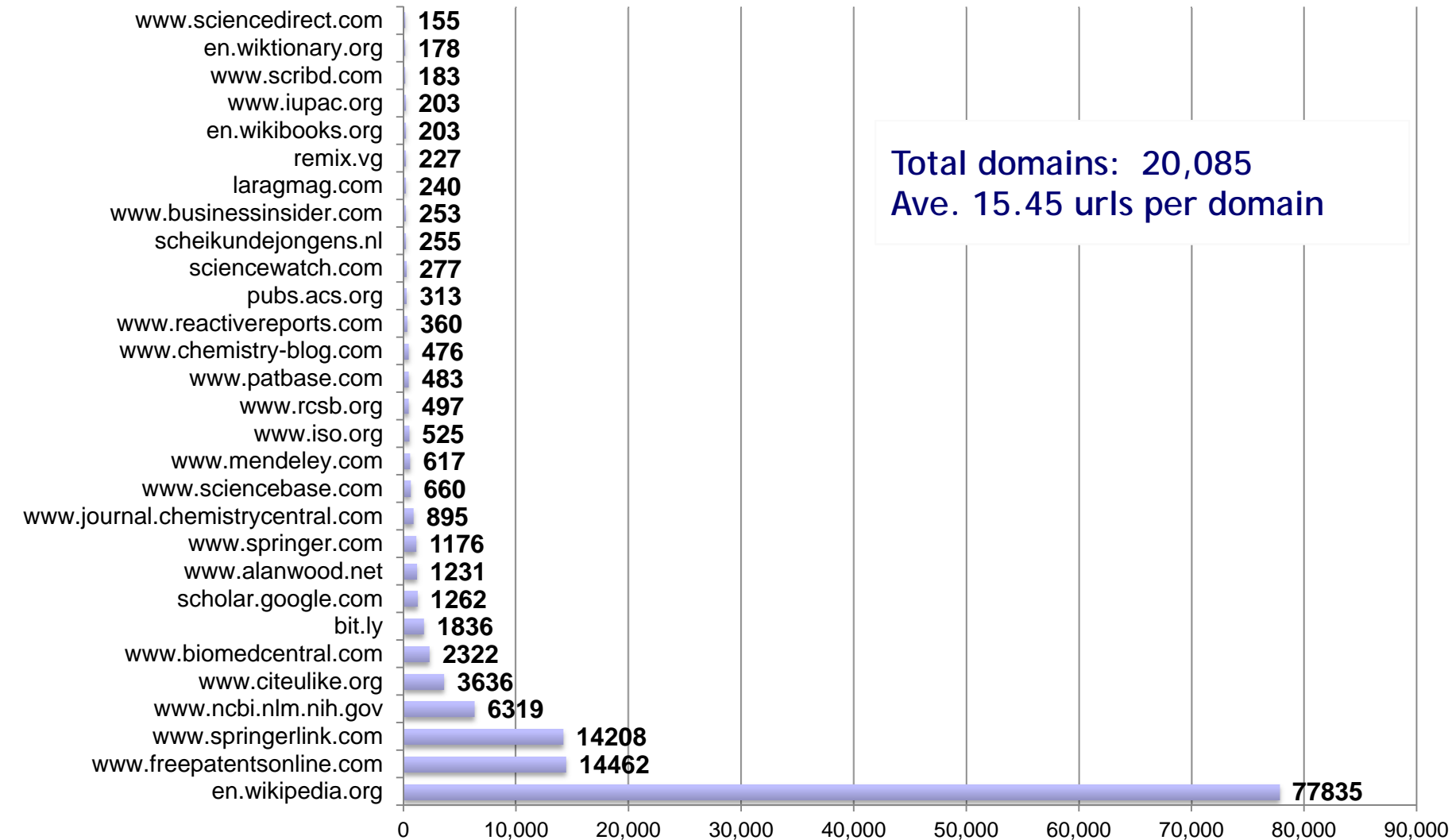
- URL's visited: 309,504
- total number of chemical names:
  - 4,336,593 (17.37 names/page)
- unique names extracted: 249,557
- structures extracted: 180,311
- unique visitors: 54,204
- avg visitors/day: 318
- avg/longest time on site: 4:34 / 28:41

# How busy are they?

Sessions / page views



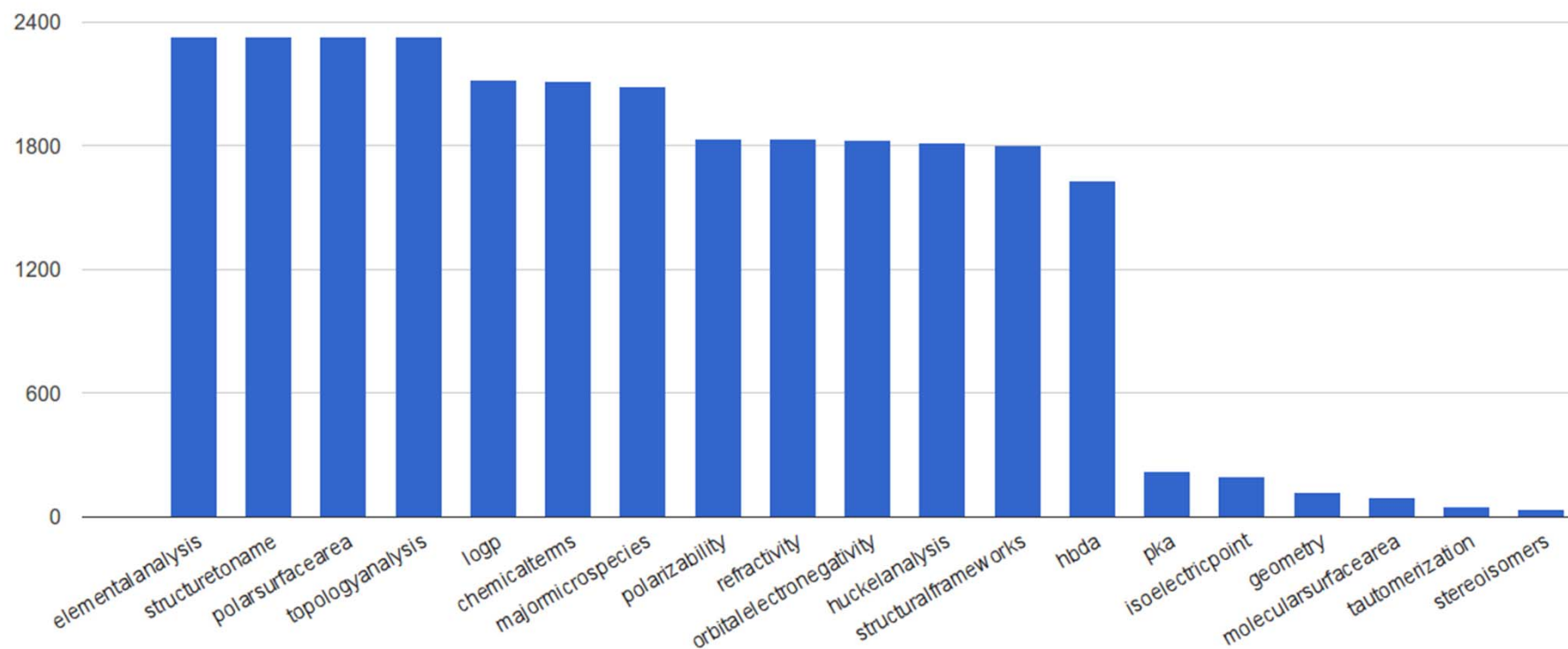
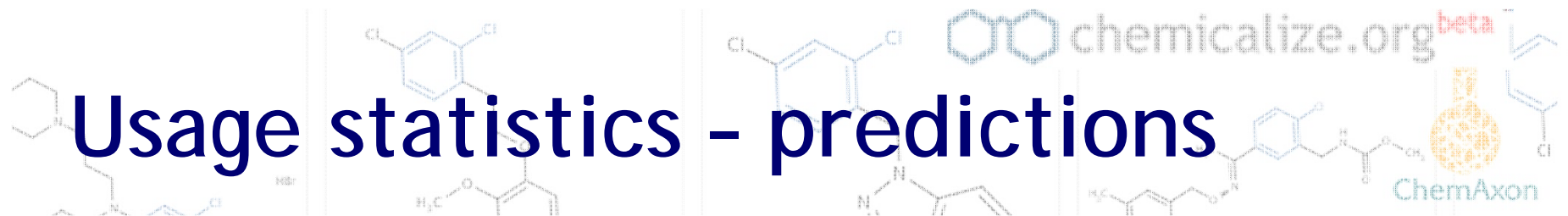
# Top domains



# Top pages

- [ncbi.nlm.nih.gov/sites/entrez?cmd=Link](http://ncbi.nlm.nih.gov/sites/entrez?cmd=Link)
- [springer.com/?SGWID=0-102-0-0-0](http://springer.com/?SGWID=0-102-0-0-0)
- [en.wikipedia.org/w/index.php?title=List\\_of\\_World\\_Health\\_Organization\\_Essential\\_Medicines](http://en.wikipedia.org/w/index.php?title=List_of_World_Health_Organization_Essential_Medicines)
- [journal.chemistrycentral.com/search/results.asp?db=pm](http://journal.chemistrycentral.com/search/results.asp?db=pm)
- [en.wikipedia.org/wiki/Aspirin](http://en.wikipedia.org/wiki/Aspirin)
- [en.wikipedia.org/wiki/List\\_of\\_organic\\_compounds](http://en.wikipedia.org/wiki/List_of_organic_compounds)
- [en.wikipedia.org/w/index.php?title=ATC\\_code\\_J01](http://en.wikipedia.org/w/index.php?title=ATC_code_J01)
- [biomedcentral.com/info/ifora/figuretypes/freepatentsonline.com/y2005/0037033.html](http://biomedcentral.com/info/ifora/figuretypes/freepatentsonline.com/y2005/0037033.html)
- [vivo.colostate.edu/hbooks/pathphys/endocrine/pancreas/insulin\\_physics.html](http://vivo.colostate.edu/hbooks/pathphys/endocrine/pancreas/insulin_physics.html)

# Usage statistics - predictions

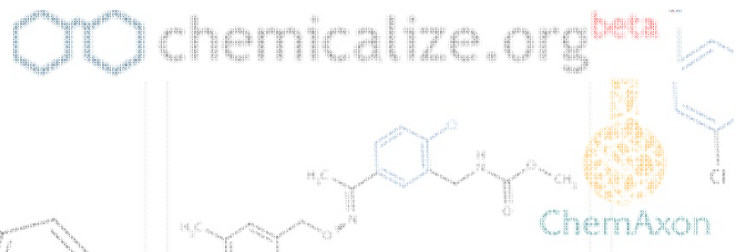
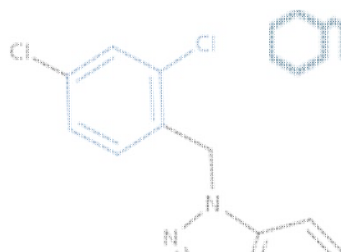
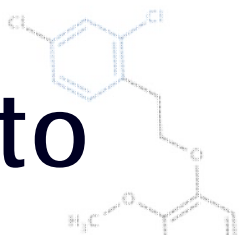


# Future plans..?

- Remaining free
- Crowdsourcing - new structures/names, bug reporting
- Working on sorting and ordering results (biggie)
- Personalization (login) = personal search history, profiles (notifications), dictionaries, calculation/search parameter settings
- Index page as window into internet chemistry use
- Text search
- Browser Plugins = chemicalize better, particularly in login/https pages (plugins tech approaching unity anyway)
- How about working up the chemistry side such as pharmacophore search, other screening, etc - there is a lot of ChemAxon tech here to play with
- Work on quality of name parsing, black lists etc
- What else guys - this is a provisional list



# Thanks to



**Andras Stracz**

Site implementation



**Daniel Bonniot**

Document & Name to structure



**Alex Allardyce, Ferenc Csizmadia**

Features, project management, idiot and advanced testing



**Zsolt Kocsmarszky**

Design



**Roland Molnar**

JChem Web Services