

Tag-Cloud Drawing: Algorithms for Cloud Visualization

Authors: Owen Kaser and Daniel Lemire

Jason Ye
University of Virginia

What is a tag-cloud?



http://st.depositphotos.com/1004032/3375/i/950/depositphotos_33754381-Software-development-concept-in-tag-cloud.jpg

Characteristics of a tag-cloud

Flickr tag-cloud

All time most popular tags

amsterdam animal animals april architecture art australia baby barcelona beach
berlin birthday black blackandwhite blue boston bw california cameraphone
camping canada canon car cat cats chicago china christmas church city
clouds color concert day dc dog dogs england europe family festival film florida
flower flowers food france friends fun garden geotagged germany girl
graduation graffiti green halloween hawaii hiking holiday home honeymoon hongkong house
india ireland island italy japan july june kids lake landscape light london losangeles
macro march may me mexico moblog mountain mountains museum music nature
new newyork newyorkcity newzealand night nikon nyc ocean paris park
party people photo portrait red river roadtrip rock rome san sanfrancisco school
scotland sea seattle show sky snow spain spring street summer sun sunset
sydney taiwan texas thailand tokyo toronto travel tree trees trip uk urban usa
vacation vancouver washington water wedding white winter yellow york
zoo

- Visual representation
- Navigator
- Histogram

Two types of tag-clouds

Inline Text

Popular Tags

AdSense Affiliate marketing Author rank Authorrank
Authorship ccTLD **CMS** Content Management System Content writing
service DC DNS Domain Name System Domains Drupal Duplicate
content Extensions Facebook **Google** Google+
communities Google AdSense Google Analytics Google Pagerank
Google **Plus** Google PR gTLD **Joomla**
Joomla 3 Joomla extensions Link building Linkbuilding
Magic quotes Make money on website Name server Outsourced content
writing Pagerank Plugins **Search** **Engine**
Optimization **SEO** Social networks Spam TLD
Twitter Webhosting Webhosts **Wordpress**

Arbitrary Placement

within people plains having company information followed
 ground principal rivers action extent except foundation together
 engineers without during during crossing eBooks charde
 elephants length running thousand including easily saccas
 degrees toward called available important distances carrying
footnote required turkestan things communication gutenberg general
 before difficult ghazni gisits quetta **country**
 district follows jelahabad ponies carried nature angloiranian states united london institution service
 should village nearly described angloiranian region comprises
 possibly caspian hundred southern central russia england
 authority occupation described khyber horses
 organization strength pounds excellent
 copyright mountainous afghan
 society association tribes
 border however feature
 russians
valley argandab defences certain animals supply routes
 direct supplies certain
 almost either organized
troops cavalry divisions transport operations
 reserve modern regiments system better present railway
 indian regular besides purposes
 regiment native mounted
 natives
 officers
 although european
 spanish
 armenian
 mountaineer
 garrison
 mountain
 artillery
 through
 cannot little natural
 guards division beyond another
 establishment
 elephant saddle across
 illustration behind passes khelbar pishin
military
 general
 forces officer political
 soldiers
 government
 roberts khan kouloun
 british
 strong
 passed persian
 territory
russian

Two types of tag-clouds

Inline Text

- Order of text has no semantic meaning
- Paragraph made exclusively from inline elements (span, em, i)
- Excessive clumps of white space

Arbitrary Placement

- Tags can be reordered, placement depends on relationships
- HTML nested tables
- Wasted space

Inline Text

Algorithm 1: Break up an ordered list of tags

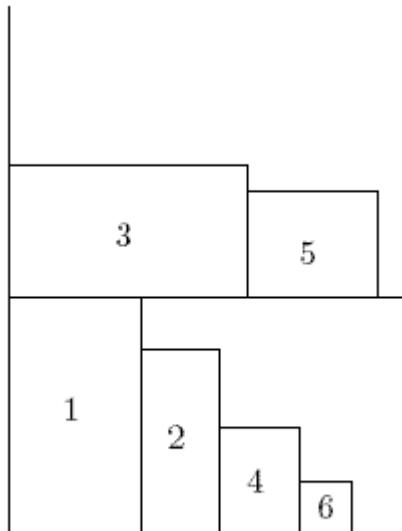
- Greedy Algorithm: $O(n)$
- Knuth-Plass Algorithm: $O(n^2)$
 - Compute *badness* of fit
 - Minimize sum of squares of each line's badness
 - Reconstruct optimal badness recursively

Inline Text

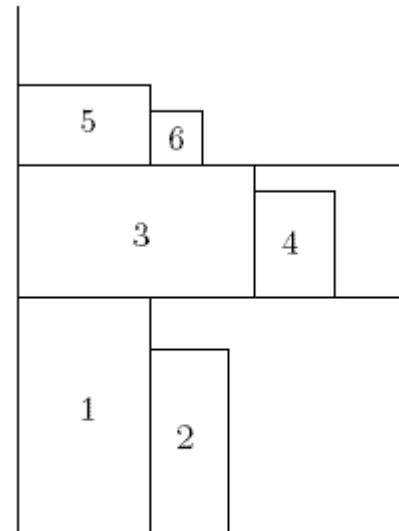
Algorithm 2: Reorders tags to decrease badness

- NP-hard Strip Packing Problem (SPP)
 - Use dynamic programming to place tags optimally while keeping the best solution
- First Fit Decreasing Height, Weight (FFDHW)

SPP Approximation Algorithms



FFDH



NFDH

<http://cgi.csc.liv.ac.uk/~epa/ffdh.GIF>

<http://cgi.csc.liv.ac.uk/~epa/nfdh.GIF>

Results: Inline text

ℓ_1 norm: the sum of all the “badness”

- FFDH and FFDHW is much better than dynamic programming

ℓ_2 norm: the sum of all the squares of “badness”

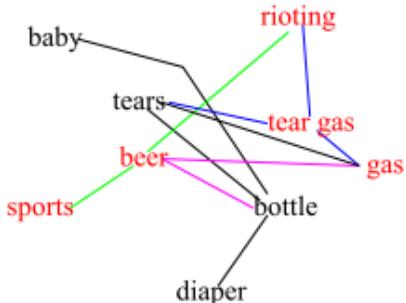
- FFDH and FFDHW only slightly better, dynamic programming is a competitive solution

Arbitrary Placement

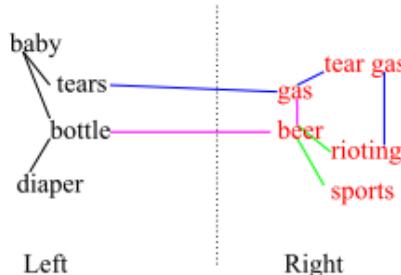
Algorithm: Electronic Design Automation (EDA)

1) Min-cut Placement: NP-hard

- Bipartitioning into “right” and “left”



(a) Bipartitioning: before

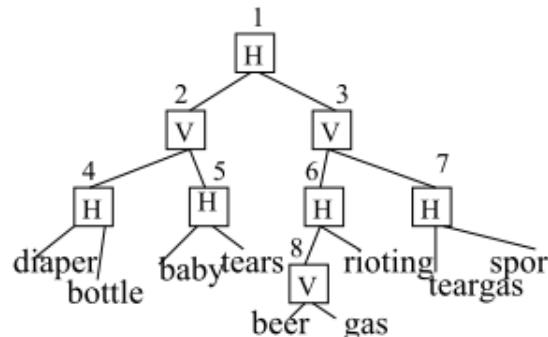


(b) Bipartitioning: after

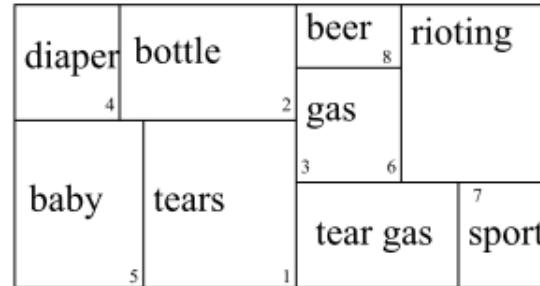
Arbitrary Placement

2) Slicing floorplans

- Recursive bipartitioning represented by slicing tree



(a) Slicing tree. Numbers next to nodes relate to areas in the slicing floorplan.



(b) Slicing floorplan

Arbitrary Placement

3) Nested Tables

- Every internal node in tree is a 2-element table

mother	little	father	pisistratus	fathers
roland		before indeed	caxton	vivian
captain	should			seemed
myself	thought	ellinor	trevanion	himself
through	without	though		

(a) Displaying table borders

mother	little	father	pisistratus	fathers
roland		before indeed	caxton	vivian
captain	should			seemed
myself	thought	ellinor	trevanion	himself
through	without	though		

(b) Displayed with appropriate CSS

```
<table><tr>
  <td> <table>
    <tr><td>beer</td></tr>
    <tr><td>gas</td></tr>
  </table></td>
  <td>rioting</td>
</tr></table>
```

Results: Arbitrary Placement

- Greedy method used 2-17% less area than min-cut
 - However, min-cut approach much better for semantic proximity