



Behind Google Instant

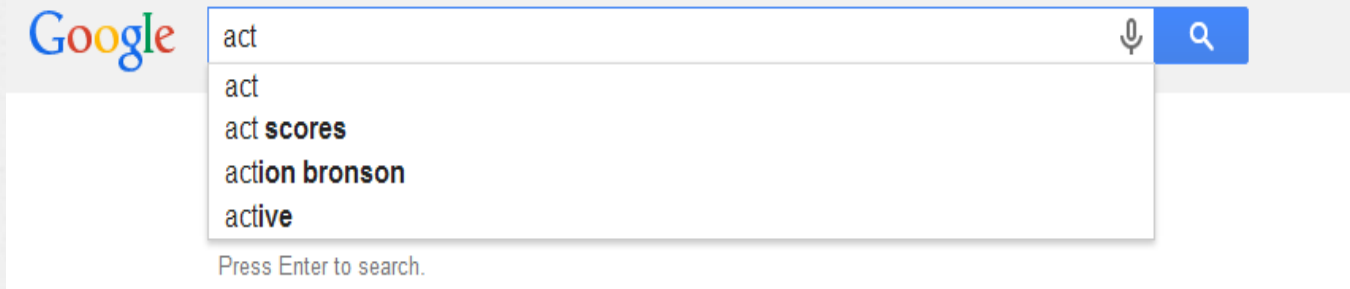
Ryan Murphy

Outline

- o Autocomplete
- o Fuzzy Search Algorithm
- o Initial Autocomplete Layout
- o How Google Differs
- o Final Autocomplete Layout
- o Possible Improvements

Autocomplete

- Return ranked results
- Uses fuzzy search



Fuzzy Search Algorithm

- Problem:

- Compare autocomplete substrings to search string

- Solution:

- Inefficient – $O(n^3m)$
- Efficient – Dr. Peter Sellers's modification to Levenshtein's distance algorithm

Substring Edit Distance

- o Two differences:
 - o Zero out first row – allows for substring comparison, $O(nm)$
 - o Keep track of calculation path – reveals substring match, $O(n+m)$

		C	A	C	T	U	S
	0	0	0	0	0	0	0
A	1	1		1	1	1	1
C	2	1	1		1	2	2
T	3	2	2	1		1	2

		C	A	C	T	U	S
A C T		0	0	0	0	0	0
	A	1	1	0	1	1	1
	C	2	1	1		1	2
	T	3	2	2	1		2

		C	A	C	T	U	S
A C T		0	0	0	0	0	0
	A	1	1	0	1	1	1
	C	2	1	1	0	1	2
	T	3	2	2	1	1	2

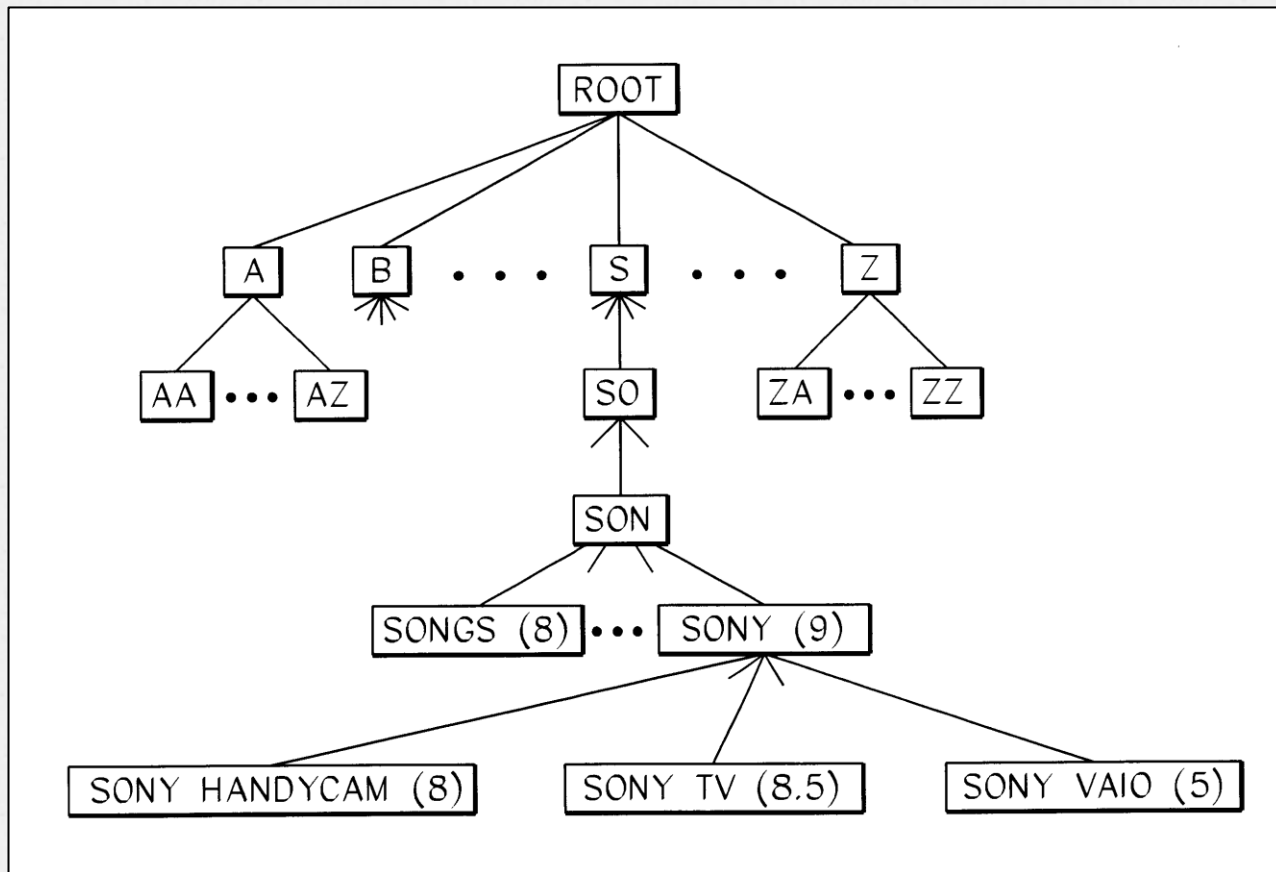
		C	A	C	T	U	S
A C T		0	0	0	0	0	0
	A	1	1	0	1	1	1
	C	2	1	1	0	1	2
	T	3	2	2	1	0	2

		C	A	C	T	U	S
	0	0	0	0	0	0	0
A	1	1	0	1	1	1	1
C	2	1	1	0	1	2	2
T	3	2	2	1	0	1	2

Initial Autocomplete Layout

- o Autocomplete dictionary
- o Gather possible autocomplete via fuzzy search
- o Filter based on rank
- o Return top X results
- o NOT exactly how Google works

Google Tree Structure



Why Fuzzy Needed?

- o Tree traversal
 - o not spelling dependent
 - o faster than fuzzy calculation
- o What if no results on first pass?

Final Autocomplete Layout

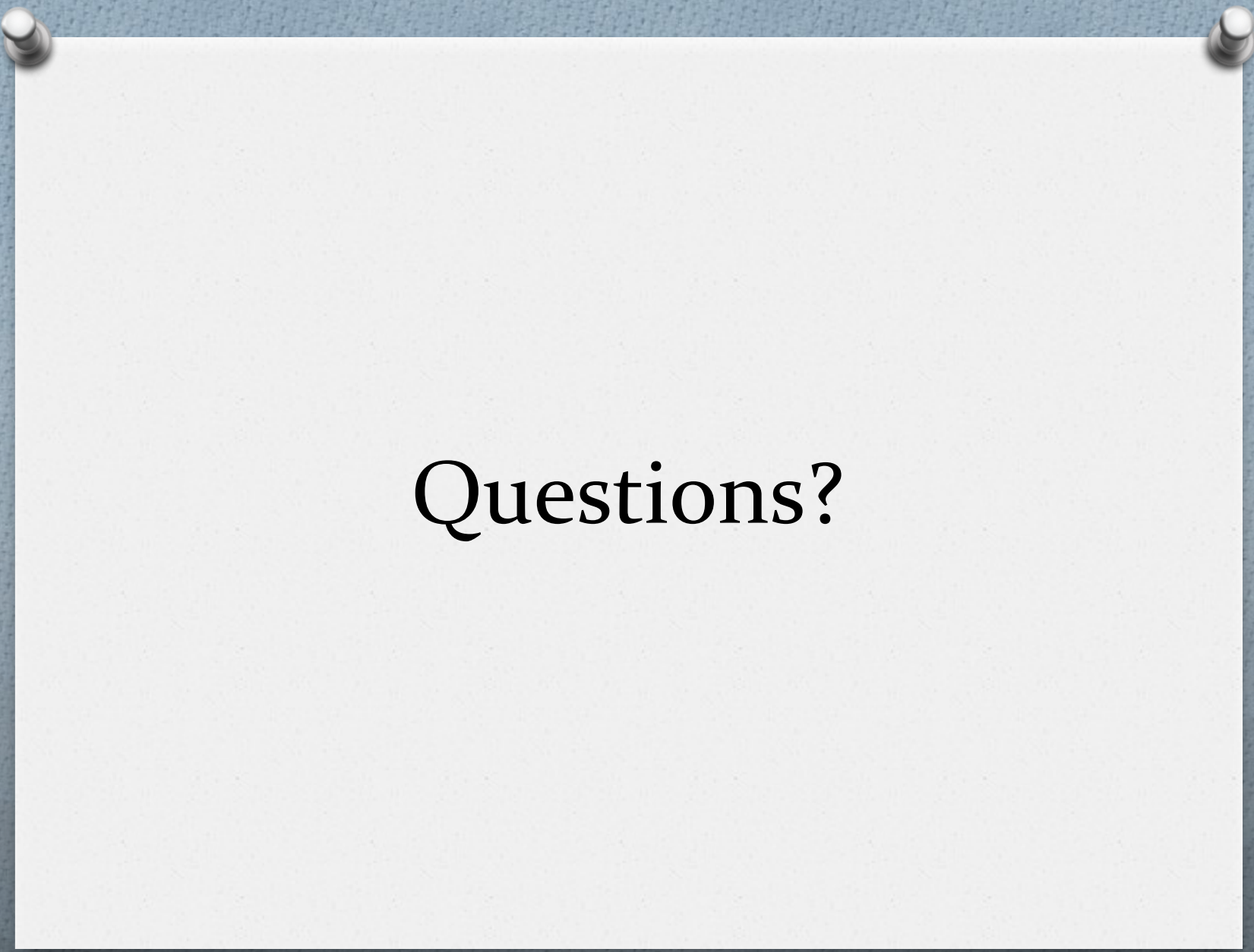
- o Create autocomplete tree
- o Gather possible autocomplete via leaf nodes
- o If nothing gathered first time:
 - o Use fuzzy search for comparison
 - o Gather possible autocomplete via leaf nodes
- o Filter based on rank
- o Return top X results

Possible Improvements

- Search-as-you-converse
- Search-as-you-read

Sources

- o "Approximate string matching." *Wikipedia*. Wikimedia Foundation, 20 Apr. 2014. Web. 9 May 2014.
<http://en.wikipedia.org/wiki/Approximate_string_matching>.
- o Ortega, Ruben, John Avery, and Robert Frederick. "Patent US6564213 - Search query autocompletion." *Google Books*. United States Patent and Trademark Office, 13 May 2003. Web. 9 May 2014.
<<http://google.com/patents/US6564213>>.
- o Yates, R., and Berthier Neto. *Modern Information Retrieval*. 2nd ed. New York: ACM Press, 2011. Print.



Questions?