

Web: separate texts

According to an eye tracking study by Cornell's L.A. Granka T. Joachims and G. Cay, the first listing at Google's SERP (results page) gets 56% of clicks, the second gets 13% and the 11th gets about 1%.

Wikipedia: synthesis

"Knuth is a fan of Wikipedia, but he's a bit leery of the concept, saying that he would not want to have to remain forever on guard after making technically complex contributions, lest his comments be badly reedited."
http://en.wikipedia.org/wiki/Donald_Knuth
 contributed by knuth-pbds15.Stanford.EDU

What's next?

Both biological evolution and modern economy are based on competition and exchange!
 If the Web is a medieval village economy and Wikipedia is a state-monopolistic system, what's next?

Lightweight distributed real-time version control for wikis

CAUSAL TREES

Victor Grishchenko, Ural State University, Russia

<http://oc-co.org>

A simple case...

Test

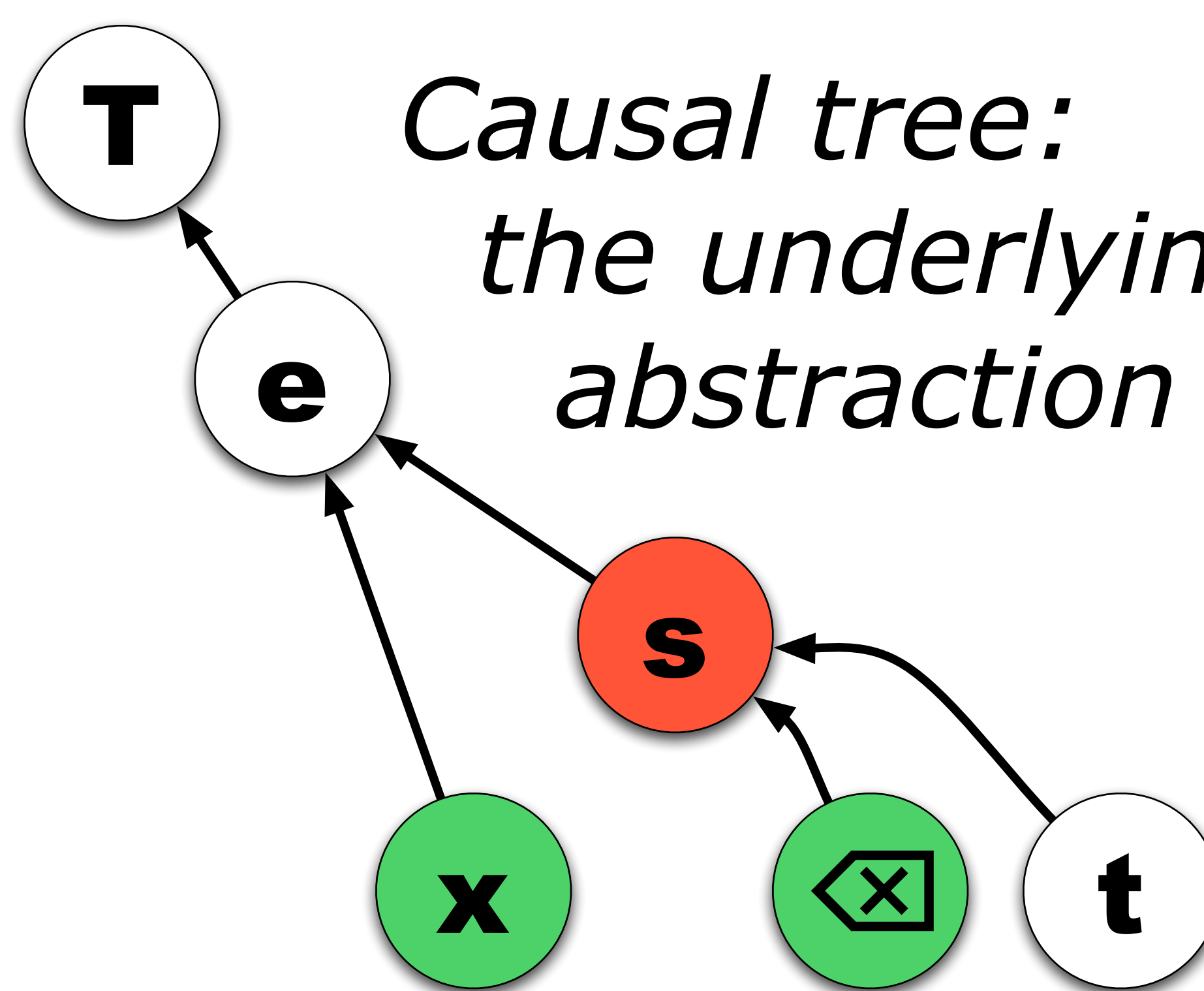
changed to...

Text

000	001	002	003	004	005	006
000	000	001	002	003	003	002
ॐ	T	e	s	t	⊗	x

Feed: append-only storage format

Causal tree:
the underlying abstraction



atom id
causing atom id
symbol

000	001	002	006	003	005	004
000	000	001	002	002	003	003
ॐ	T	e	x	s	⊗	t

Weave: version control data structure

mutation and crossover for texts!

BASICS

- * text consists of "atoms"
- * atom has unique id
- * there is a *causing* atom (i.e. originally preceding)

Atoms form a tree; linear form (the text) is defined as depth-first preorder traversal.

ALGORITHMS

Text assembly and other basic weave algorithms operate in terms of arrays and indexes, do single or multiple passes: O(N) cheap.
 Use "advanced" data structures (hashtables) for importing new atoms only.

STORAGE

Storage space: plain text times 20 (x10 for non-Latin).
 Append-only feed is easily updated and easily checked for updates - trivial caching and sync.

THE ATOM TYPE

```

struct atom_t {
  uint64_t id;
  uint64_t cause;
  wchar_t symbol;
}
  
```

BENEFITS

- Real-time changes
- Per-symbol granularity
- Per-second timestamps
- Per-symbol authorship
- Easy caching
- Easy version merge
- Convergence (OT sense)
- Instant versioning
- Lighter than a VCS