

# xenodot

XML management

**start**

let's reverse engineer

Google™



**what's known?**

Backlinks!

The image shows a screenshot of a Google search result page. At the top, the Google logo is visible on the left, and the search bar contains the word "failure". To the right of the search bar are links for "Web" and "Bilder". Below the search bar, there are options for "Suche: Das Web", "Seiten auf Deutsch", and "Seiten aus Deutschland". A blue bar below the search bar indicates "Web" and "Ergebnisse 1 - 50 von ungefähr 510.000.000 für failure. (0,68 Sekunden)". The first search result is titled "Biography of President George W. Bush" with a link to "www.whitehouse.gov/president/gwbbio.html" and 27k backlinks. The second result is titled "Biography of Jimmy Carter" with a link to "www.whitehouse.gov/history/presidents/jc39.html" and 36k backlinks. The third result is titled "Google verbindet Bush mit 'Miserable failure ...'" with a link to "www.ringfahndung.de/index.shtml" and 11k backlinks. The status bar at the bottom of the browser window shows "Done".

Google

Web Bilder

failure

Suche

Erweiterte Suche  
Einstellungen

Suche: Das Web Seiten auf Deutsch Seiten aus Deutschland

**Web** Ergebnisse 1 - 50 von ungefähr 510.000.000 für failure. (0,68 Sekunden)

[Biography of President George W. Bush](#) - [ [Diese Seite übersetzen](#) ]  
Biography of the 43rd President of the United States.  
[www.whitehouse.gov/president/gwbbio.html](http://www.whitehouse.gov/president/gwbbio.html) - 27k - [Im Cache](#) - [Ähnliche Seiten](#)

[Biography of Jimmy Carter](#) - [ [Diese Seite übersetzen](#) ]  
Short biography from the official White House site.  
[www.whitehouse.gov/history/presidents/jc39.html](http://www.whitehouse.gov/history/presidents/jc39.html) - 36k - [Im Cache](#) - [Ähnliche Seiten](#)

[:: ringfahndung :: Google verbindet Bush mit 'Miserable failure ...](#)  
Google verbindet Bush mit 'Miserable failure', also kläglichem versagen! \*\*\*. Dies mag damit zusammenhängen, dass Google langsam auch ein moralisches und ...  
[www.ringfahndung.de/index.shtml](http://www.ringfahndung.de/index.shtml) - 11k - [Im Cache](#) - [Ähnliche Seiten](#)

Done



Google Directory - Kids and Teens > School Time > Social Studies > History > By Region > N...

# Google Directory

Google Search | Directory Help

Search only in Bush, George Walker

## Bush, George Walker

[Kids and Teens](#) > [School Time](#) > [Social Studies](#) > [History](#) > [North America](#) > [United States](#) > [Presidents](#) > Bush, George Walker

Web Pages	Viewing in Google PageRank order	View in alphabetical order
<a href="#">George W. Bush</a>	<a href="http://www.whitehouse.gov/president/gwbbio.html">http://www.whitehouse.gov/president/gwbbio.html</a>	
Biography of the president from the official White House web site.		
<a href="#">Online NewsHour: Inauguration 2001</a>	<a href="http://www.pbs.org/newshour/inauguration/index.html">http://www.pbs.org/newshour/inauguration/index.html</a>	
A PBS guide to the inauguration of America's 43rd President. Includes the full Inaugural Address, a look at Bush's cabinet, photos, and classroom resources.		
<a href="#">George W. Bush: Like Presidential Father, Like Presidential Son</a>	<a href="http://www.infoplease.com/ce6/people/A0809584.html">http://www.infoplease.com/ce6/people/A0809584.html</a>	
Beth Rowen compares the father-son presidential figures in a brief article.		
<a href="#">The Life of George W.</a>		

Done

Page rank!

http://www.google.de - XML - Google-Suche Anmelden

Google   [Erweiterte Suche](#)  
[Einstellungen](#)

Suche:  Das Web  Seiten auf Deutsch  Seiten aus Deutschland

---

**Web** Ergebnisse 1 - 50 von ungefähr 2.200.000.000 für XML. (0,19 Sekunden)

[Extensible Markup Language - Wikipedia](#)  
XML definiert dabei die Regeln für den Aufbau solcher XML-Dokumente, welches wohlgeformt ist und ein durch eine Grammatik definiertes Dokument ist.  
[de.wikipedia.org/wiki/XML](http://de.wikipedia.org/wiki/XML) - 48k - [Im Cache](#) - [Ähnliche Seiten](#)

[Ajax \(Programmierung\) - Wikipedia](#)  
(In diesem Kontext wird vornehmlich XML als Trägermedium der einzelnen Daten ...  
Viele Ajax-Toolkits versenden XML-Daten, indem sie die in diesem Bereich ...  
[de.wikipedia.org/wiki/Ajax\\_\(Programmierung\)](http://de.wikipedia.org/wiki/Ajax_(Programmierung)) - 69k - [Im Cache](#) - [Ähnliche Seiten](#)

[xml-web / XML](#)  
Deutschsprachige Seiten zu den Themen XML, XSLT, Formatting Objects, CSS und SVG.  
[www.xml-web.de/](http://www.xml-web.de/) - 6k - [Im Cache](#) - [Ähnliche Seiten](#)

[cenShare Redaktionssystem](#)  
Redaktionssystem, CMS und Workflow für effizientes Publishing mit XML  
[www.coware.de](http://www.coware.de)

[XML/XHTML schnell lernen](#)  
Diese Investition in Ihr Wissen lohnt, weil jahrelang breit nutzbar  
[www.linuxhotel.de](http://www.linuxhotel.de)

[InfoPath und XML](#)  
Integration in Office

Done

Word indexing!



update frequency

everything?

probably!

# requirements

arbitrary access to content **and** structure

➔ loss-less storage

aggregation and search across documents

➔ full-indexing

access to changes over time

➔ versioning

random content (e.g. by crawler)

➔ no structural assumptions possible

**sounds like an xmldb!**

plenty too choose from!?

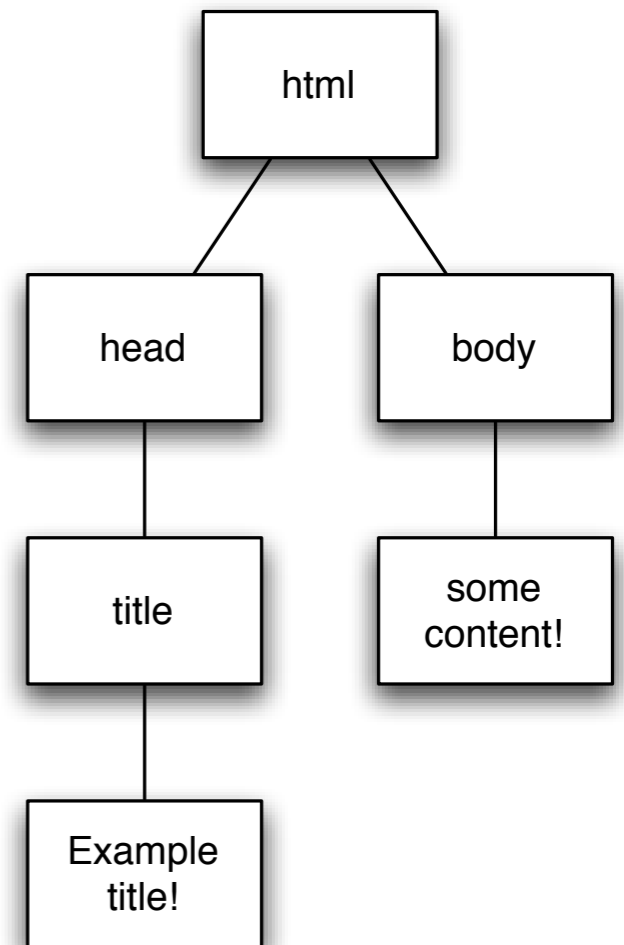
**none suitable!**

**two questions!**

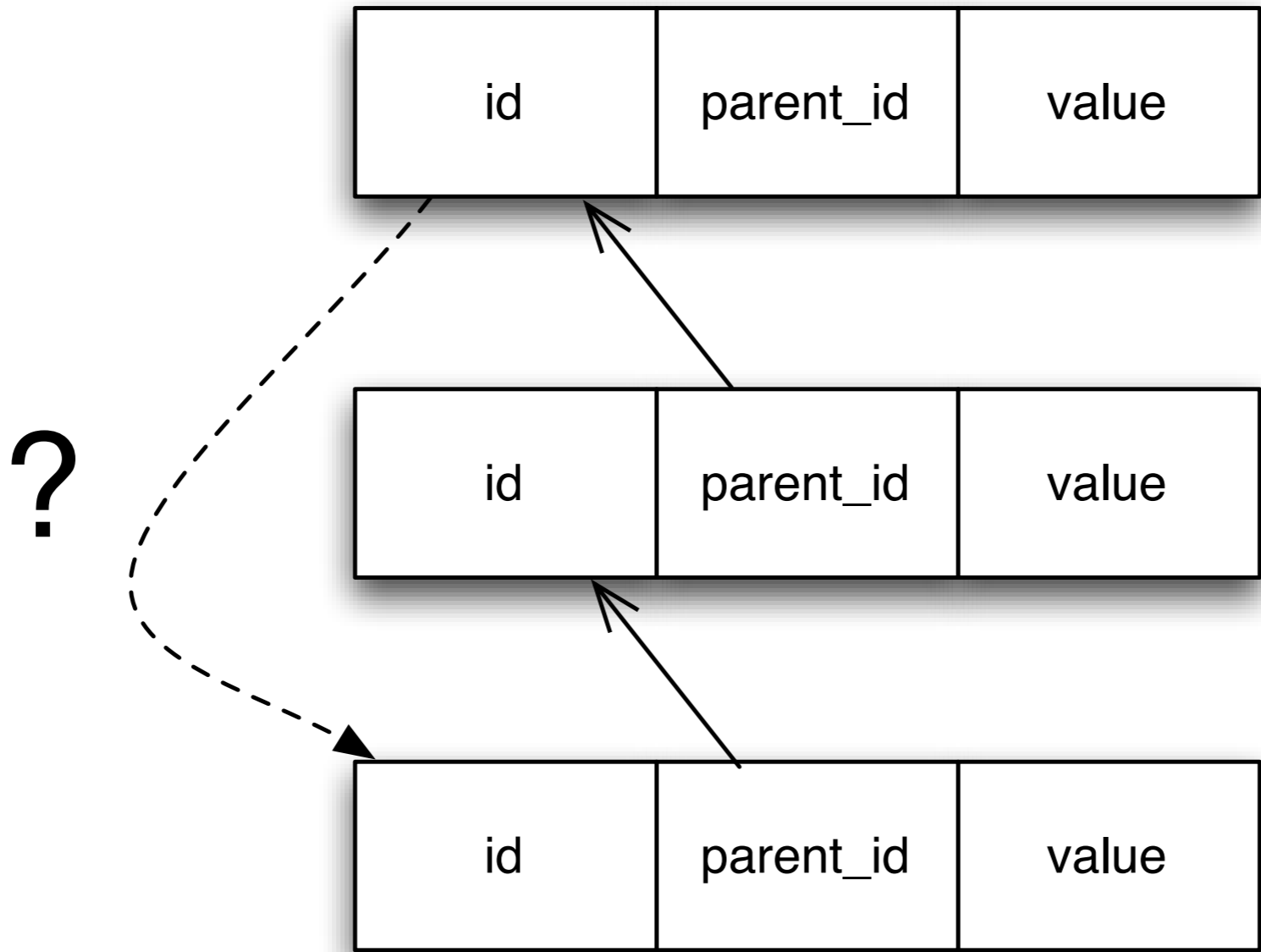
**what to store?**

**how to access?**

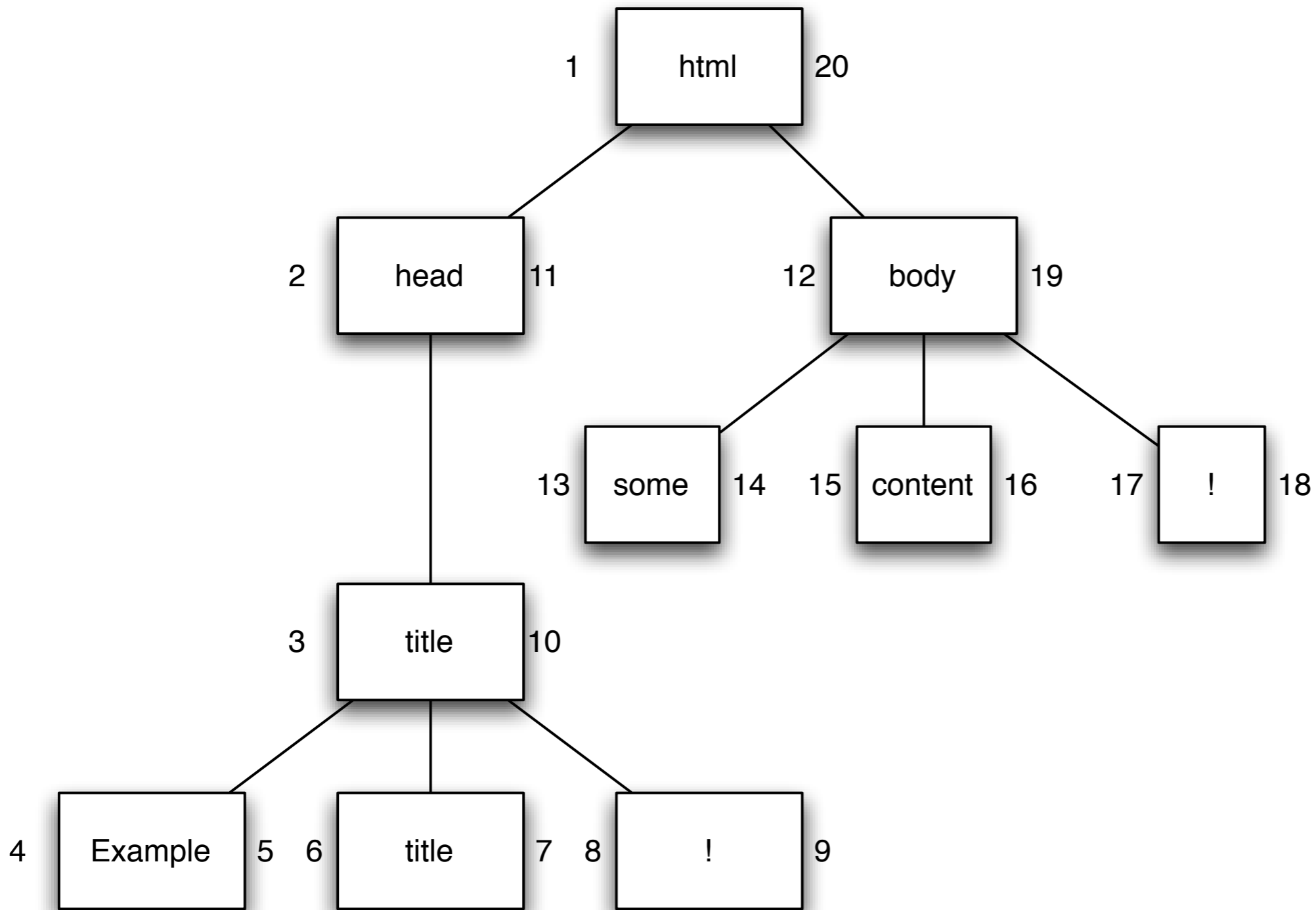
```
<html>
  <head>
    <title>Example title!</title>
  </head>
  <body>
    some content!
  </body>
</html>
```



# the problem of hierarchy



# nested sets

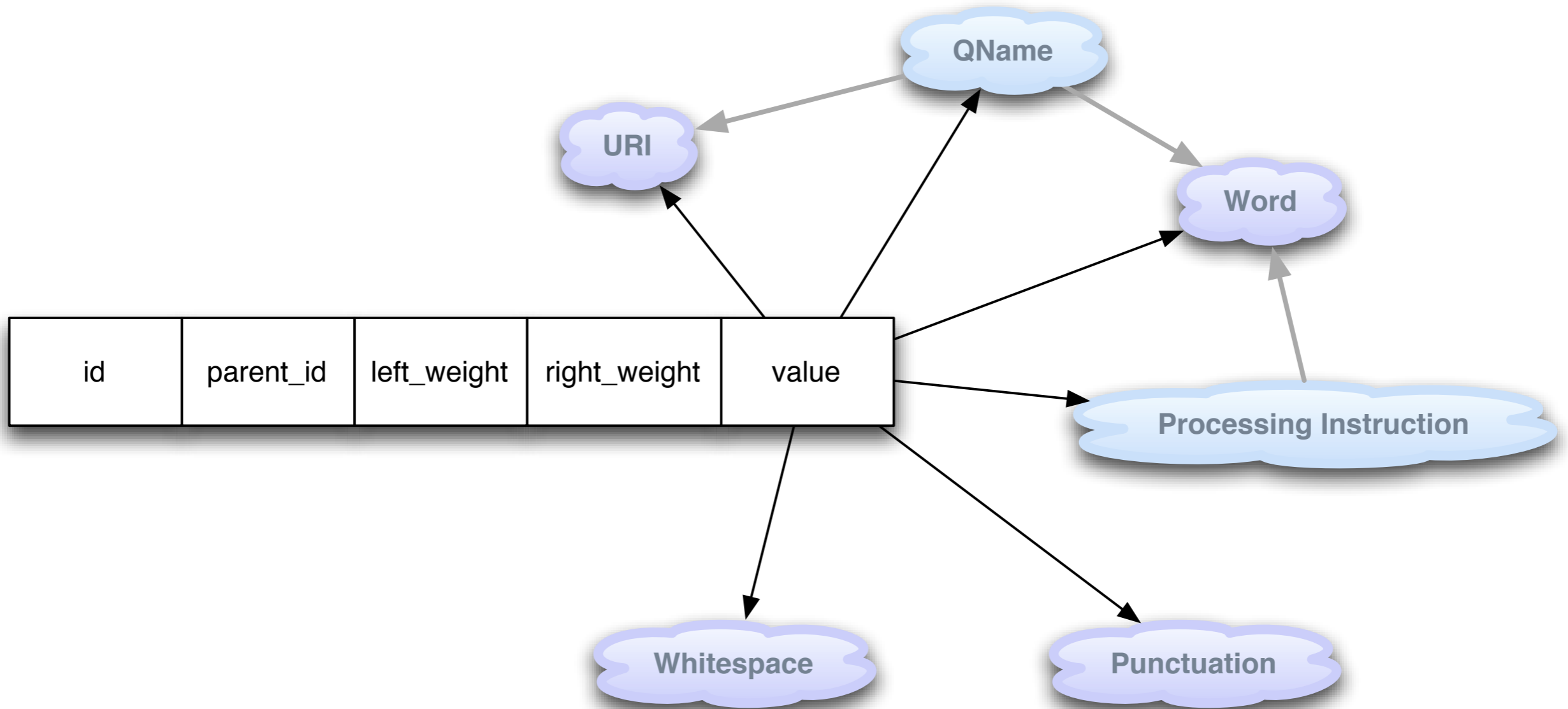


id	parent_id	left_weight	right_weight	value
----	-----------	-------------	--------------	-------

```

select ... where left_weight > curr.left_weight
and right_weight < curr.right_weight
  
```

value?



**polymorphic foreign keys!**

value\_type

id	table
200	processing_instruction
2100	qname
4100	word

id	parent_id	left_weight	right_weight	value_type	value_id
----	-----------	-------------	--------------	------------	----------

word

id	table
1	Example
2	title
3	some

full indexing?

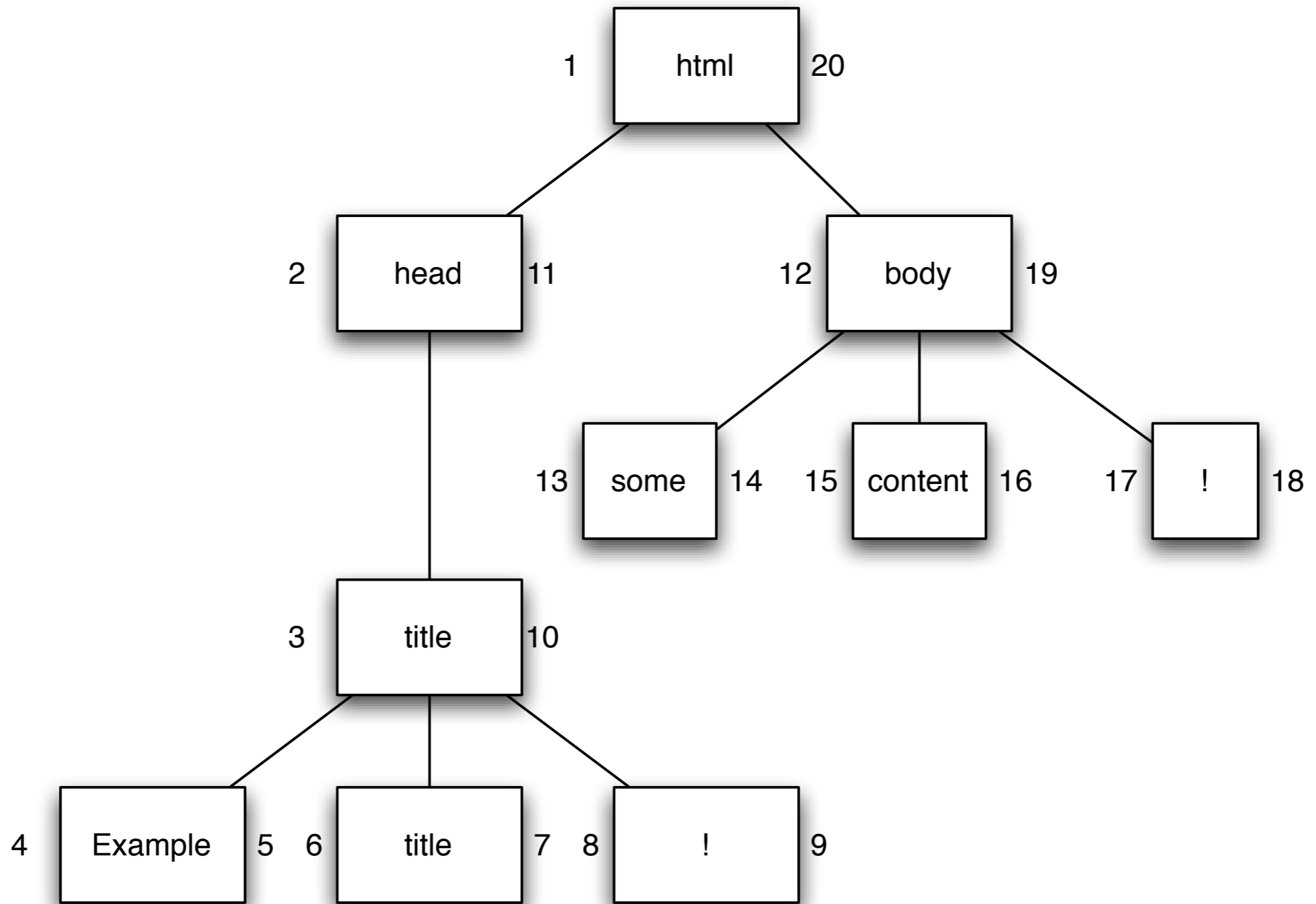
**no text nodes**

but

word

**whitespace**

**punctuation**



on to question 2:  
how to access?

**first**

**code is written in ruby**

**runs on the client**

runs in the db (pl/ruby)

**uses higher-order messages**

higher what?

**very complicated way**

**to say**

**Xenodot[:all].html.a!.content!.to\_s**

**produces only one sql-statement**

demo!

**current work**

Xenodot[:all].html.a!.where{|a| a[:href] == 'http://target.com'}.content!

non-ruby api

**vision**

**be the backend for  
next-generation apps**

**thank you for you attention!**

**questions? feedback?**

- Zenodot became the first superintendent of the Library of Alexandria (284 BC)
- He is the inventor of the *textual criticism*, a method of comparing several versions of a document to reconstruct the archetype
- Xenodot is a versioning xmldb optimized for search