

3/06/19

Google

"Not Gone; But Forgotten — Helping users
refind [lost] web pages by identifying those
that are likely to be lost"

Lost Rank → Rank results in refinding search engines
according to likelihood of their being lost
to the user

hygi → temporal document access patterns
to determine documents that are both
important and not recently accessed.

(4) → understanding user's topical access
patterns to determine topics that are
more unfamiliar — diff of refinding

→ our ability to formulate queries to
locate specific documents decays over time

→ As time grows, doc. coll grows ↑↑
and good querying ability ↓↓

→ Page Rank, HTTP promote credible (4)
authenticating pages | lots of i/c links from
other credible pages

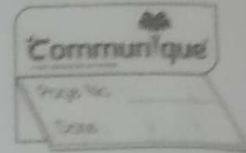
→ lost info → which is hidden within a
difficult Navigation path / file deeply nested dir.

Ranking Components:-

Accum pattern - time of last accu
2 Pages accrued 1 month ago. } equal lot rank.

P_1 \wedge P_2 (one/month over 2 yrs)
(viewed @ time) prob (loss) = ↓
weaker lot rank

pg vary freq accessed and then
a long list \rightarrow dormant pages.



\Rightarrow dor pages = important but lacks familiarity due to

⊗ Topic patterns \rightarrow
topic revisiting varies across topics.

eg: & for code doc \rightarrow Navigational style.

than link for health doc \rightarrow explorative style.

\rightarrow Build an access pattern for each topic
associate topical activity to each page.

"pages with more dormant topics - \uparrow to be lost

{ Java class thing - specific = higher lost rank
Java " - generic = weaker lost rank;

⑧ Difficulty before original accn:-

→ path to reach the original page blocked

✓ → specific queries - navigational style - easy to find

→ pages offer long trails of queries and

✓ query reformulations — indicate explorative queries

→ user recollection of terms from pages from different task will be used

energy efficient buildings → passive windows
(not keep them)

∴ likely to be lost