Revista da UIIPS – Unidade de Investigação do Instituto Politécnico de Santarém,
Vol. 2017, N. º 2, 2017, pp. 105-116
ISBN: 2182-9608
http://ojs.ipsantarem.pt/index.php/REVUIIPS

STIGMERGIC HYPERLINK'S CONTRIBUTES TO WEB SEARCH

Artur Marques

Escola Superior de Gestão e Tecnologia, Instituto Politécnico de Santarém, Portugal

artur.marques@esg.ipsantarem.pt

ABSTRACT

Stigmergic hyperlinks are hyperlinks with a "heart beat": if used they stay healthy and online; if neglected, they fade, eventually getting replaced. Their life attribute is a relative usage measure that regular hyperlinks do not provide, hence PageRank-like measures have historically been well informed about the structure of webs of documents, but unaware of what users effectively do with the links.

This paper elaborates on how to input the users' perspective into Google's original, structure centric, PageRank metric. The discussion then bridges to the Deep Web, some search challenges, and how stigmergic hyperlinks could help decentralize the search experience, facilitating user generated search solutions and supporting new related business models.

Keywords: hyperlinks, PageRank, social search, stigmergy, www

1. INTRODUCTION

An individual "stigmergic hyperlink", or "stigh", is a WWW hyperlink with a floating vitality attribute that captures what users have been doing with it (Marques & Figueiredo, 2010b): the higher the vitality, the greater its relative usage.

A plural system of stighs exhibits an automatic recommendation system type of behavior, via a Nature-inspired form of indirect communication named "stigmergy" (Grassé, 1959). This behavior drives some interesting applications, namely an approach to the dead-links problem, but here I focus on a couple of WWW search related contributes:

- a usage informed measure of PageRank (Page, Brin, Motwani, & Winograd, 1998), that addresses the absence of hyperlinks usage data in the original computation;
- stighs as vehicles for search decentralization, via custom search solutions, that could be Deep Web (Bergman, 2001) oriented.

The search engine Google sorts its results using, among other inputs, a "PageRank" (PR) metric that tries to measure the voting that is going on, at the scale of the World Wide Web, for a certain resource R: all pages P_n that link to R, are casting votes on R. The more the votes, the better placed the resource will probably be in a search results list. The original PageRank is structure-based, not informed of what users have been doing with the voting resources – this happens because regular hyperlinks do not provide any usage data. In contrast, stigmergic hyperlinks do inform about their relative usage and so provide a way for vote casting to be weighted by effective followers, instead of considering all pointers equal.

The Deep Web is the Web beyond the reach of conventional search engines; it poses a search problem that is being addressed with horizontal scalable solutions that can't deal with certain barriers, and with vertical non-scalable tools that will only work for very specific contents. Stighs automatically provide a "search the destination" functionality that authors can rewrite. This means that in the hypothetical scenario of their pervasiveness, they would not only be supporting a search decentralization mechanism, but also nurturing a context that would facilitate custom specialized solutions.

First I introduce key concepts: stigmergy, hyperlinks and stigmergic hyperlinks. Next I elaborate on the intended contributes. I conclude with some final remarks.

2. CONCEPTS

2.1 Stigmergy and the WWW

The word stigmergy comes from the Greek "stigma" (mark) and "ergon" (work), meaning "the mark of work". The expression is credited to Pierre-Paul Grassé, a French biologist who studied social insects and observed collective behaviors beyond the capabilities of individuals alone and without a centralized coordination. In particular, Grassé studied termites while constructing a nest (Grassé, 1959) and noticed that such a relatively complex structure is the result of simple interactions of individuals with their immediate surroundings, not requiring direct communication between termites.

Termites aggregate terrain flocks or mud balls while building their nest. Which mud balls and where they are to be placed depends on pheromones, i.e. depends on the concentration of certain marks that evaporate over time. The pace of evaporation and the pheromones' intensity play to give rise to different structures, such as columns or arches. There is no plan and no individual has a global memory of the terrain, nor that is necessary to accomplish the building tasks: it suffices to know the immediate surroundings.

Prior to Grassé, but without explaining the termites' "invisible communication", Eugène Marais also wrote on the subject (Marais, 1937) and made analogies with the human body.

My analogy, or the analogy in "stigmergic hyperlinks" – to be elaborated on section 2.4 –, is that a web page is like a terrain being worked by its users/visitors, who will be leaving marks on it whenever clicking hyperlinks; such marking is a form of indirect communication to others; it translates to a relative usage mechanism that also models evaporation over time. The WWW is an example of a shared medium that people can write to and read from.

Search engines, such as Google, consider hyperlinks a form of mark/citation and a metric named PageRank tries to assess the "importance" of resources from the link structure of the WWW (Brin & Page, 1998; Page et al., 1998). PageRank feeds on the link structure of the WWW but it also indirectly influences it in a self-reinforcing positive feedback fashion, because top/bottom ranked resources are more/less likely to become pointed by others (Eysenbach & Köhler, 2002; Yue, Patel, & Roehrig, 2010).

Stigmergy, as a form of indirect communication, is effective for some distributed control problems (Dorigo, Birattari, & Stützle, 2006) and it constitutes an approach where one can focus on relatively simple concepts that are at the core of more complex structures and behaviors that will emerge, as long as the basilar elements are properly modeled.

2.2 Hyperlinks

Hyperlinks are the core of hypertext (Wardrip-Fruin, 2004), connecting parts of a same document (internal links) and documents to other documents (external links). The World Wide Web can be considered a single hypertext system. Hyperlinks were devised as means for non-linear content consumption, but they can be extended and play a wider role.

"Hyperlinks are more than technical artefacts" (Maeyer, 2013) supporting a gamut of situations in a way which could eventually be different if other attributes and behaviors were also available, for example if using stigmergic hyperlinks.

2.3 Hyperlinks in search

The Google search engine was among the first to incorporate hyperlinks in the computation of search results (Brin & Page, 1998) – hyperlinks pointing to a destination are interpreted as recommending the destination. Google differentiated itself by computing the link structure of the Web and by considering a hyperlink's anchor text as information about its destination object and not so much about the source. The graph that represents the link structure of the Web is obtained by crawling regular hyperlinks and it feeds the "PageRank" (PR) algorithm that helps in sorting the search results.

Subjacent to Google's PageRank is a "random surfer" user behavior model: someone who is browsing pages, clicking on hyperlinks until bored; then he will pick a random page, for example using the browser's bookmarks. The higher a page's PageRank, the higher the probability the random surfer will land there. The probability that the random user will stop clicking and request another location by other means, is called the "damping factor" (d) and is usually set to 0.85, as an estimate of how regularly browser functionalities such as the bookmarks are used.

Google ranks its search results using, among other inputs, the mentioned "PageRank" (PR) metric (Page et al., 1998) that tries to measure the voting that is going on, at the scale of the WWW, for a certain resource R.

PageRank is a "way to attach a score to web pages on the basis of their connectivity" (Bianchini, Gori, & Scarselli, 2005) and it intentionally ignores all contents except for the hyperlinks, as one way to avoid "search engine persuasion" or "web spamming" techniques that usually affect information retrieval scoring algorithms, that take into account page contents (Marchiori, 1997).

All pages P_n that link to a resource R, are considered to be casting votes on R. But a P_n page eventually also links to other resources, so its voting power is equally divided by them all: it is this undifferentiated division that could change, using stighs.

The more the votes R gets, the better placed the resource will probably be in a search results list. The original PageRank is structure-based and usage challenged, since it does not input what users have been doing with the voting hyperlinks – this happens because regular hyperlinks do not provide any usage data.

Each page P_n pointing to a page R has its say on R's vote, originally expressed by

$$PR(P_n) \times \frac{1}{\#LinksGoingOutOf\,(P_n)}$$

The expression accounts all external links in P_n as equally probable destinations; however, if the external links were stighs, each could provide a reading of its own vitality; hence it would be possible to assess the visitors' relative preferences and discriminate among the external hyperlinks.

The contribution that stigmergic hyperlinks could have to the computation of PageRank and customized variants (Tsoi, Hagenbuchner, & Scarselli, 2006), relates to the division of the available PR(P_n) vote: knowing which hyperlinks are most used for exiting P_n, creates conditions to distribute the vote in function of the visitors' preferences. This is detailed in section 3.1.

2.4 Stigmergic hyperlinks

Regular Web hyperlinks have limitations, namely unidirectional linkage, unverified destination and issues related to relevance, reputation and trust (Leuf, 2006), some addressed by technologies like W3C's XLink (W3C, DeRose, Maler, Orchard, & Walsh, 2010).

Stigmergic **h**yperlinks – "stighs" for short – were designed to automate the replacement of neglected or undesired destinations in a set of hyperlinks, and so to present a solution to the dead or broken links problem – an enduring phenomenon, acknowledged as important since the early Web days to the present (Ashman, 2000; Davis, 1998; Kobayashi & Takeda, 2000; Kovilakath & Kumar, 2012; Le, 2013; Markwell & Brooks, 2002; Martinez-Romo & Araujo, 2012) –, following a stigmergic approach that achieves a recommendation system type of behavior and fits (Parunak, 2005)'s architecture and taxonomy for stigmergy.

Stighs can look and behave like regular hypertext hyperlinks, but they have a life attribute and run at the server side (Marques & Figueiredo, 2010b). Their life is reinforced on every click and fades over time as the result of "natural" decay: every few page clicks, every stigh in the page gets weaker, eventually until "death" by replacement with an alternative sourced from the survivors in the same system – the algorithm for the replacement process is detailed in (Marques & Figueiredo, 2010b). Internally, the life attribute is a number and how much it increases or decreases, in response to attention or to decay, is definable – these and other configurable values influence the pace of events in a system of stigmergic hyperlinks. Such configurability is by design, intending to facilitate experiments and research.

One thing is to understand, to define, and to classify broken links; another is to identify and/or assist in fixing them; another yet is to have self-sustained automatic responses to the issue. Only two projects, other than stighs, are found explicitly attacking the problem of automatically fixing broken links: "DSNotify" (Haslhofer & Popitsch, 2009; Popitsch & Haslhofer, 2010) and "WISH" (Morishima, Nakamizo, Iida, Sugimoto, & Kitagawa, 2009).

"DSNotify" was born as a solution to detect broken links in LOD (Linking Open Data) sources: in that context a broken link occurs when a request can't obtain the solicited resource description (RDF), because the target was moved or removed. For removed targets, the data source can act by eliminating all statements containing the removed link target; for moved targets it is necessary to find the same resource and update the link target: the main challenge lies in answering if what happened was a move or a remove.

The LOD context of the "DSNotify" project makes it very different from the stigmergic hyperlinks' raw WWW environment. "DSNotify" is a "A2A" (application-to-application) approach because it is conceived as an assistant to applications who want to preserve link integrity in their own data sets, or to be notified of changes in data sources.

"WISH", or "Web Integrity management by Self-Healing mechanisms", is a "project for the development of software tools to help maintain the integrity of Web content". It works on the "regular" WWW, so it is more comparable to the stigmergic hyperlinks.

"WISH" provides "LIM" (Link Integrity Management) tools: functions to identify broken links, moved web pages and to where they move. One significant difference to stighs is the "audience" for the technology: "WISH" appears to aim at hyperlink curators: actors who want to preserve the structural integrity in collections of regular hyperlinks. An actor first registers the URLs he wants to monitor using the "LIM" (Link Integrity Manager) tool, which will then start monitoring the pointers; if and when there is a disruption, a "PageChaser" component goes chasing for the new location of the resource, using the corresponding and previously stored contents as a guide. "PageChaser" usually returns more than one possible new location, so a choice must be made. The selection process involves a custom ranking "LA" or "Link Authority server". "LA" supplies "link authorities" which are "well maintained" and "up-to-date" link containers that, because of their quality, should be helpful in spotting the new address for the moved resource.

The Stigmergic hyperlinks are for web authors who want to have hyperlink populations that adapt to a community's input, providing the extra value of automatically eliminating the neglected destinations and some other extra features.

Stighs abstract the reason why the vitality of hyperlinks fluctuates the way it does, so when a replacement happens it is irrelevant if it eliminates a broken link due to deletion, or a broken link due to moving, or a working link to contents with drastic semantic changes, or just a relatively underappreciated working link.

The way stighs compare to other approaches for its core "broken links solution" feature is not the focus of this paper, which instead decides to depart from the core to the periphery and question what could eventually happen to Web search, if stighs became disseminated. This dissemination has not happened and that is an event out of the control of any individual alone, yet some hopefully interesting theoretical considerations can be made, as discussed next in section 3.

3. STIGHS’ CONTRIBUTES TO WEB SEARCH

3.1 A usage informed PageRank

Stighs keep record of what users have been doing with them. Data is gathered implicitly and anonymously, meaning that on every click some client/server data is automatically logged, including the stighs’ life, providing a reading of relative usage that can assist PageRank-like metrics in becoming aware of the user’s behavior.

Since conventional hyperlinks do not provide methods to directly assess their usage, search results tend to be structure centric, biased to the content production side of the WWW: links are seen as their authors’ votes, not their users’. In a scenario of generalized availability of hyperlinks’ usage data, this bias could be addressed.

Relatively new media, like Twitter, convey the same structure vs. network usage unbalance, with the structural "following" and "followed" measurements telling a different story from the one told by the number of "retweets". As an analogy to the Web, the "following" are the outgoing links from a page to other pages (votes on others), the "followers" are like incoming links (votes from others), and "retweets" are the exact diffusion of someone else’s contents to all that are "following".

While "following" and "followed" are mainly structural, retweets are an explicit usage of the network for content redistribution, and it seems sensible to consider them a "stronger" vote than just "following".

It was measured that ranking by retweets does not match ranking by followers (Kwak, Lee, Park, & Moon, 2010), with some that do not lead on "followers", leading on retweets. For example, during the period the authors studied the entire Twitter network, the actor Ashton Kutcher ranked first by number of followers and by PageRank based on the followers/followed structure, but "only" 13th by number of retweets, while Pete Cashmore, not listed in at least the top 20 most followed, was #1 by retweets.

By providing relative usage data, stighs support PageRank-like variants that incorporate the usage/consumption side of the WWW, down to the hyperlink granularity.

Here is the general expression for the relative usage of a stigmergic hyperlink x:

$$relativeUsage(stigh_x) = \frac{vitality(stigh_x)}{\sum_{i=1}^{numberOfStighsInThePage} vitality(stigh_i)}$$

Notice that

$$\sum_{i=1}^{numberOfStighsInThePage} relativeUsage(stigh_i) = 1$$

As previously discussed about PageRank, each page P_n pointing to a page R will have its say on R's vote, originally expressed by and worth

$$PR(P_n) \times \frac{1}{\#LinksGoingOutOf(P_n)}$$

This expression accounts all external links in P_n as equally probable destinations, simply dividing P_n 's vote by the number of possible exits.

If the external links were stighs, it would be possible to do the voting in proportion to their effective usage.

For example, imagine that page P_n has three stighs {s1, s2, s3}, respectively linking to pages {P1, P2, P3}, with "lives" {L1=150, L2=25, L3=25} – in this case, the sum of all "lives" is 150+25+25=200 and the relative browsing preferences are

{ relativeUsage(s1) = 150/200=3/4, relativeUsage(s2) = 25/200=1/8, relativeUsage (s3) = 25/200=1/8 }.

Users are six times more probable to leave the page via s1, than via the other links, so when voting on P1, does it make more sense to dilute P_n's vote equally by 3, or by a different amount that reflects the effective preferences?

The alternative to the "all links are equal" perspective is to factor P_n's vote on a page P_x , by the relative usage of the outgoing link that in P_n points to P_x:

$$\text{amount of vote of } P_n \text{ on } P_x = PR(P_n) \times relativeUsage(stigh_x) \text{ , } stigh_x \text{ being a pointer from } P_n \text{ to } P_x$$

So, for the example above, assuming Q as the quantity or voting amount P_n has to give, the usage informed vote of P_n on P1 is worth Q*(3/4), because s1 linking to P1 has a relative usage of 3/4. If all links are considered equal, the vote on P1 is only worth Q*(1/3), i.e. the same that are worth the votes on P2 and P3, despite those destinations being less used.

In conclusion, the expression for an alternative PageRank for a page R, $PR_{alt}(R)$, based on stighs' relative usage data is the sum of all the votes contributed by all the pages which have a link to R, using relative usage information instead of the default non-discriminating egalitarian division: each page P_n's vote on R is its PageRank not simply divided by its number of outgoing links, but instead factored by the relative usage of the stigh which in P_n links to R.

$$PR_{alt}(R) = \sum_{p=1}^{\text{\# pages pointing to R}} PR(Page_p) \times relativeUsage(\text{stigh in } Page_p \text{ linking to } R)$$

3.2 WWW search challenges

The "Deep Web" expression is credited to Michael Bergman (Bergman, 2001), referring to the part of the WWW that search engines "cannot see". Contents might escape search engines' efforts, because

- They explicit request to be ignored;
- They are computed dynamically, from the interaction with the user, and don't have a static representation that can be stored and/or searched;
- They don't have a dedicated URL and thus cannot be indexed by systems that, by design, need that URL to exist – this affects major search engines;
- The trend towards a more personalized WWW might require more computation and less open databases.

Databases hiding behind HTML forms are considered a significant part of the "Deep Web", addressed in efforts that try to pre-compute the corresponding URL requests, as in Google's "surfacing" (Madhavan et al., 2008). Every search expression that a user inputs into a GET search form has its equivalent URL and causes a server reply. As an alternative to brute force data retrieval mechanisms that could request entire dictionaries, techniques that identify sets of keywords sufficient to extract most of, or the entirety of the "hiding" databases, have been developed (Barbosa & Freire, 2004). Once these URLs are available, conventional search engines can search their respective contents and point to the results, since there is a unique corresponding URL.

However, the pre-computation of URLs is only possible for GET forms, not for POST forms. The main challenge with POST forms is that, no matter how different the client's requests and the server's responses, the exposed URL is always the same, impeding indexing by address.

The size of the Deep Web has been estimated to be hundreds of times bigger than the "surface" Web (He, Patel, Zhang, & Chang, 2007) and the causes that already contributed to this scale remain active, namely the usage of Web databases, dynamic content generation and personalization.

Horizontal approaches to the Deep Web solve general problems, such as URLs extraction from GET forms, and provide results that have already been incorporated into commercial search engines, including Google (Madhavan et al., 2008), but much content remains out of the reach of umbrella solutions. On the other side of the spectrum there are specialized tools: the directory at http://completeplanet.com claims to index 70,000 vertical search engines.

3.3 Addressing search challenges, one stigh at a time

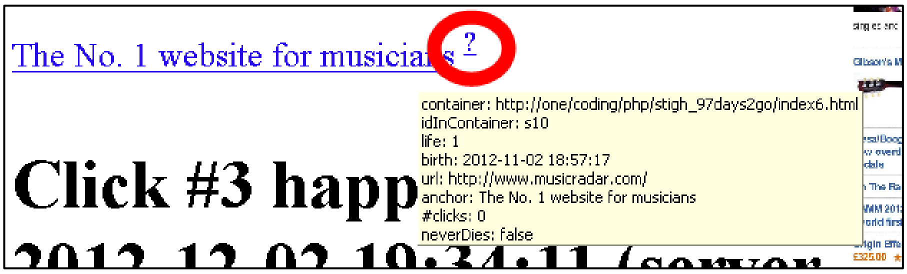
Stigmergic Hyperlinks expose a search-the-destination method: the idea is to go beyond linking and provide a contextualized search experience.

The stigh's current search functionality does the following…

- Links to a custom search solution, specific for the destination;
- Automatically writes a simple HTML search interface, specific for the pointed URL, that will accept a search expression and call the site specific search solution;
- Automatically codes a meta-search solution over the destination. This default solution is a "meta search" because it blends results that Google, Yahoo and Bing would produce for the query/destination pair. The results are direct pointers, stripped from trackers and without any filtering, so they might appeal to privacy conscious users;
- Makes it easy for authors to edit the HTML and redesign the search interface and/or edit the PHP and rewrite the searching itself, paving the road for original solutions, eventually

specialized to the point of being able to fetch results from "deep webs" that the major search engines, by definition, can't handle.

When enabled, the search function will appear as a link with the "?" anchor, at the upper right corner of the corresponding stigmergic hyperlink – see Figure 1 for an example.



The No. 1 website for musicians ?

Click #3 happ

2012 12 02 19:34:11 (server

```
container: http://one/coding/php/stigh_97days2go/index6.html
idInContainer: s10
life: 1
birth: 2012-11-02 18:57:17
url: http://www.musicradar.com/
anchor: The No. 1 website for musicians
#clicks: 0
neverDies: false
```

Figure 1 : The circle highlights a stigh's link to its custom search tool.

If one clicks the question mark, the search interface will appear – Figure 2 exemplifies it with a query for "knopfler".

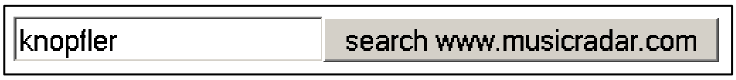
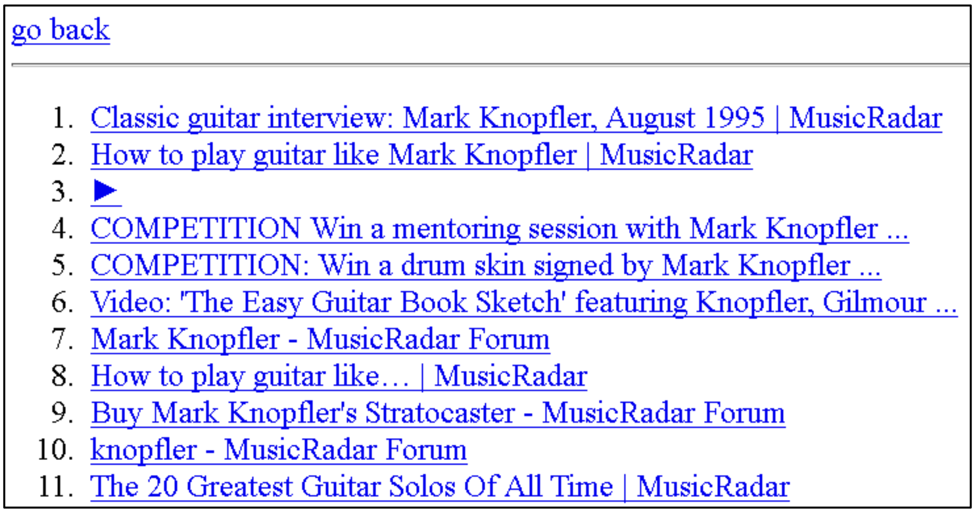


Figure 2 : The default custom search interface.

One hundred results from that site alone would appear, not all illustrated in Figure 3.



go back

1. Classic guitar interview: Mark Knopfler, August 1995 | MusicRadar
2. How to play guitar like Mark Knopfler | MusicRadar
3. ▶
4. COMPETITION Win a mentoring session with Mark Knopfler ...
5. COMPETITION: Win a drum skin signed by Mark Knopfler ...
6. Video: 'The Easy Guitar Book Sketch' featuring Knopfler, Gilmour ...
7. Mark Knopfler - MusicRadar Forum
8. How to play guitar like… | MusicRadar
9. Buy Mark Knopfler's Stratocaster - MusicRadar Forum
10. knopfler - MusicRadar Forum
11. The 20 Greatest Guitar Solos Of All Time | MusicRadar

Figure 3 : Some search results from the automatic search solution.

Notice that the only thing that a content author has to do, to access these behaviors – automatic hyperlink replacement in the event of relative neglecting, page-as-recommender-system, and search-over-the-destination –, is to declare the stigh.

The first Stigmergic Hyperlinks prototype was written in C# and could only exist in .NET pages. More recently, after rewriting to PHP, it became possible to declare stighs in any plain HTML page, using the syntax exemplified next.

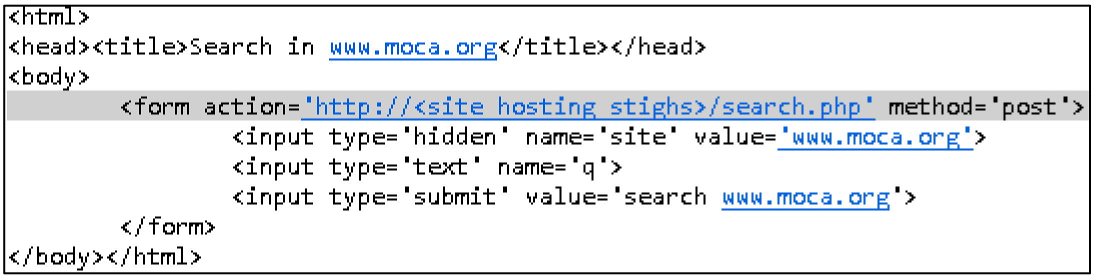
<stigh id="s1" url="http://moca.org/" life="8" anchor="The Museum of Contemporary Art"></stigh>

The container page should include the "stigh.js" Javascript file

<script src="stigh.js"></script>

and the server should be running PHP and MySQL.

The default and automatic search solutions that the stighs provide are simple meta-searches. For more elaborate search experiences, including Deep Web searches, it is necessary to edit the automatically generated search interface / HTML form, as the one exemplified next in Figure 4,



```
<html>
<head><title>Search in www.moca.org</title></head>
<body>
        <form action='http://<site hosting stighs>/search.php' method='post'>
                <input type='hidden' name='site' value='www.moca.org'>
                <input type='text' name='q'>
                <input type='submit' value='search www.moca.org'>
        </form>
</body></html>
```

Figure 4 : The automatically generated HTML for the default search interface.

and point its action to a custom resource, by replacing the reference to

http://<site hosting stighs>/search.php

with a reference to the alternative search script.

In (Marques & Figueiredo, 2010a) one such tool is discussed, to illustrate how a highly specialized solution enabled searching one particular deep web: in early 2010 all the magazines available at the biggest Portuguese digital magazines store – assineja.pt – were made searchable.

The motivations for someone to develop highly specialized search tools, can be any of the identified in studies about what drives software developers to work for free, for example in open source projects (Hars & Ou, 2002; Lakhani & Wolf, 2007; Scacchi, 2005); but there could also be money at play: content owners could pay for search solutions tailored to their consumers' specific needs, and large scale search providers could pay for the right to use third party solutions and abstract themselves from such vertical tasks.

In this decentralized search information market, I identify three main economic agents: solution providers, large scale search providers, and content providers. Table 1 captures some potential rewards and challenges of theirs.

Stigmergic hyperlinks' main role in this decentralized search experience scenario is as a mere placeholders for the search functionality, which is a separate challenge. This would be a new form of social search.

Table 1

Decentralized search market economic agents' rewards and challenges

Market agent type	Potential rewards	Eventual problems
Content provider	Owned content becomes more visible and should attract new consumers. Examples of monetization: content reselling, selling access subscriptions, serving ads.	Increased exposure can highlight security and/or capacity limitations. Increased content scrutiny can lead to litigation, if the content becomes disputed.
Large scale search engine	Increased use of custom searches using own database. Abstraction from specific sites and non-scalable approaches, delegated on other providers. Competition and redundancy of providers could contribute to better quality results.	Integration difficulties with the outsourced results.
Solution provider	The solution can be monetized.	Content owners might charge for access to contents. A dedicated solution can fail at any time if something relevant enough changes on the content's side. The solution provider and the content owner should articulate, but that will not always happen. Risk of litigation with other agents in the market.

4. FINAL REMARKS

I discussed two potential contributes from stigmergic hyperlinks to Web search: a usage-informed PageRank measure and hyperlinks as anchors for custom search solutions, which could act as an infrastructure for search decentralization. Although these contributes are not the key feature in stighs (the automatic replacement of broken links is) they are an interesting collateral result of their engineering.

The merited success of big search engines makes it hard to think of search differently. Many users wrongly perceive that "everything is being found", while some relevant resources aren’t being properly acknowledged, POST forms remain out of reach, and the "Deep Web" keeps growing.

Regarding the usage informed PageRank-like metric, its feasibility is totally dependent on the widespread presence of stigmergic hyperlinks or equivalent relative usage reading devices. Only with such omnipresence would be possible to evenly embed the users' perspective into the voting that is always happening on the Web and thus support new information ranking systems. New challenges would arise: a page's structure is relatively stable data, at least if compared to the effervescence of its hyperlinks' usage, meaning that ranking systems could become more volatile on their results, with unclear consequences on related business models and end-user service perception.

Regarding hyperlinks as anchors for specialized search services, it is a possibility less dependent on wide dissemination – there are already some vertical search engines available, enabling very specific searches. The nuance in stighs is that the most established Hypertext object, the hyperlink, could start offering customizable search solutions over its linked destination. This would increase the number of search interfaces and create conditions for changes in the entire search sphere, from search behavior to search monetization. From the end-user's point-of-view, it would open the opportunity for in-context searching prior to browsing; from the search service side, it could create opportunities for more, eventually better, search tools. Most of the time, developers probably don't want to code a solution for the resource they are linking to, and that is fine because the search function can be disabled or offered in an automatic meta fashion, just making use of the results available from the big search engines; but, in some situations, the site might be incompatible with generic horizontal searches, or the results they deliver might be of inferior quality compared to carefully designed alternatives – and those are the situations where quality could greatly improve, because the conventional approaches are simply not doing it.

The message is that it is possible to rethink Web search, from internal details to the high level search experience itself, with the backing of new business models.

5. REFERENCES

Ashman, H. (2000). Electronic document addressing: dealing with change. *ACM Comput. Surv., 32*(3), 201-212. doi: 10.1145/367701.367702

Barbosa, L., & Freire, J. (2004). *Siphoning Hidden-Web Data through Keyword-Based Interfaces*. Paper presented at the Symposium on Databases.

Bergman, M. K. (2001). The deep web: Surfacing hidden value. *Journal of Electronic Publishing, 7*(1), 07-01.

Bianchini, M., Gori, M., & Scarselli, F. (2005). Inside PageRank. *ACM Trans. Internet Technol., 5*(1), 92-128. doi: 10.1145/1052934.1052938

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30*(1-7), 107-117.

Davis, H. C. (1998). *Referential integrity of links in open hypermedia systems*. Paper presented at the Proceedings of the ninth ACM conference on Hypertext and hypermedia : links, objects, time and space---structure in hypermedia systems: links, objects, time and space---structure in hypermedia systems, Pittsburgh, Pennsylvania, USA.

Dorigo, M., Birattari, M., & Stützle, T. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine, 1*(4), 28-39.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Bmj, 324*(7337), 573.

Grassé, P. P. (1959). La reconstruction du nid et les coordinations interindividuelles chez bellicositermes natalensis et cubitermes sp. la théorie de la stigmergie: Essai d'interprétation du comportement des termites constructeurs. *Insectes Sociaux, 6*(1), 41-80.

Hars, A., & Ou, S. (2002). *Working for free? Motivations of participating in open source projects*.

Haslhofer, B., & Popitsch, N. (2009). *DSNotify - Detecting and Fixing Broken Links in Linked Data Sets*. Paper presented at the Proceedings of the 2009 20th International Workshop on Database and Expert Systems Application.

He, B., Patel, M., Zhang, Z., & Chang, K. C.-C. (2007). Accessing the deep web: a survey. *Communications of the ACM, 50*(5), 94-101.

Kobayashi, M., & Takeda, K. (2000). Information retrieval on the web. *ACM Comput. Surv., 32*(2), 144-173. doi: 10.1145/358923.358934

Kovilakath, V. P., & Kumar, S. D. M. (2012). *Semantic broken link detection using structured tagging scheme*. Paper presented at the Proceedings of the International Conference on Advances in Computing, Communications and Informatics, Chennai, India.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). *What is Twitter, a social network or a news media?* Paper presented at the WWW 2010.

Lakhani, K. R., & Wolf, R. G. (2007). Why hackers do what they do: Understanding motivation and effort in free/open source software projects. In J. Feller, B. Fitzgerald, S. A. Hissam & K. R. Lakhani (Eds.), *Perspectives on free and open source software*: MIT Press.

Le, T.-D. (2013). Learning resources in federated environments: a broken link checker based on URL similarity. *Int. J. Metadata Semant. Ontologies, 8*(1), 3-12. doi: 10.1504/ijmso.2013.054183

Leuf, B. (2006). Enhancing the Web *The Semantic Web - Crafting Infrastructure for Agency* (pp. 3-30): Wiley.

Madhavan, J., Ko, D., Kot, L., Ganapathy, V., Rasmussen, A., & Halevy, A. (2008). Google's deep web crawl. *Proceedings of the VLDB Endowment, 1*(2), 1241-1252.

Maeyer, J. D. (2013). Towards a hyperlinked society: A critical review of link studies. *New Media & Society, 15*(5), 737-751. doi: 10.1177/1461444812462851

Marais, E. N. (1937). The soul of the white ant. Retrieved 2010-11-22, 2010, from http://www.soilandhealth.org/03sov/0302hsted/030213marais/The%20Soul%20of%20the%20White%20Ant%20-%20Marais%20-%20ToC.htm

Marchiori, M. (1997). *The quest for correct information on the Web: hyper search engines*. Paper presented at the Selected papers from the sixth international conference on World Wide Web, Santa Clara, California, USA.

Markwell, J., & Brooks, D. (2002). Broken Links: The Ephemeral Nature of Educational WWW Hyperlinks. *Journal of Science Education and Technology, 11*(2), 105-108. doi: 10.1023/A:1014627511641

Marques, A., & Figueiredo, J. d. (2010a). *An approach to decentralizing search, using stigmergic hyperlinks.* Paper presented at the CENTERIS, Viana do Castelo, Portugal.

Marques, A., & Figueiredo, J. d. (2010b). Stigmergic Hyperlink: a new social web object. *IJISSC - International Journal of Information Systems and Social Change, 2*(4), 31-43. doi: 10.4018/jissc.2011100103

Martinez-Romo, J., & Araujo, L. (2012). Updating broken web links: An automatic recommendation system. *Inf. Process. Manage., 48*(2), 183-203. doi: 10.1016/j.ipm.2011.03.006

Morishima, A., Nakamizo, A., Iida, T., Sugimoto, S., & Kitagawa, H. (2009). *Bringing your dead links back to life: a comprehensive approach and lessons learned*. Paper presented at the Proceedings of the 20th ACM conference on Hypertext and hypermedia, Torino, Italy.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. http://infolab.stanford.edu/~backrub/pageranksub.ps

Parunak, H. V. D. (2005). Expert assessment of human-human stigmergy: Altarum Institute.

Popitsch, N. P., & Haslhofer, B. (2010). *DSNotify: handling broken links in the web of data.* Paper presented at the WWW.

Scacchi, W. (2005). Free and open source development practices in the game community. *Software, IEEE, 21*(1), 59-66.

Tsoi, A. C., Hagenbuchner, M., & Scarselli, F. (2006). Computing customized page ranks. *ACM Trans. Internet Technol., 6*(4), 381-414. doi: 10.1145/1183463.1183466

W3C, DeRose, S., Maler, E., Orchard, D., & Walsh, N. (2010). XML Linking Language (XLink). http://www.w3.org/TR/xlink11/

Wardrip-Fruin, N. (2004). *What hypertext is.* Paper presented at the ACM Conference on Hypertext.

Yue, Y., Patel, R., & Roehrig, H. (2010). *Beyond position bias: examining result attractiveness as a source of presentation bias in clickthrough data.* Paper presented at the WWW.