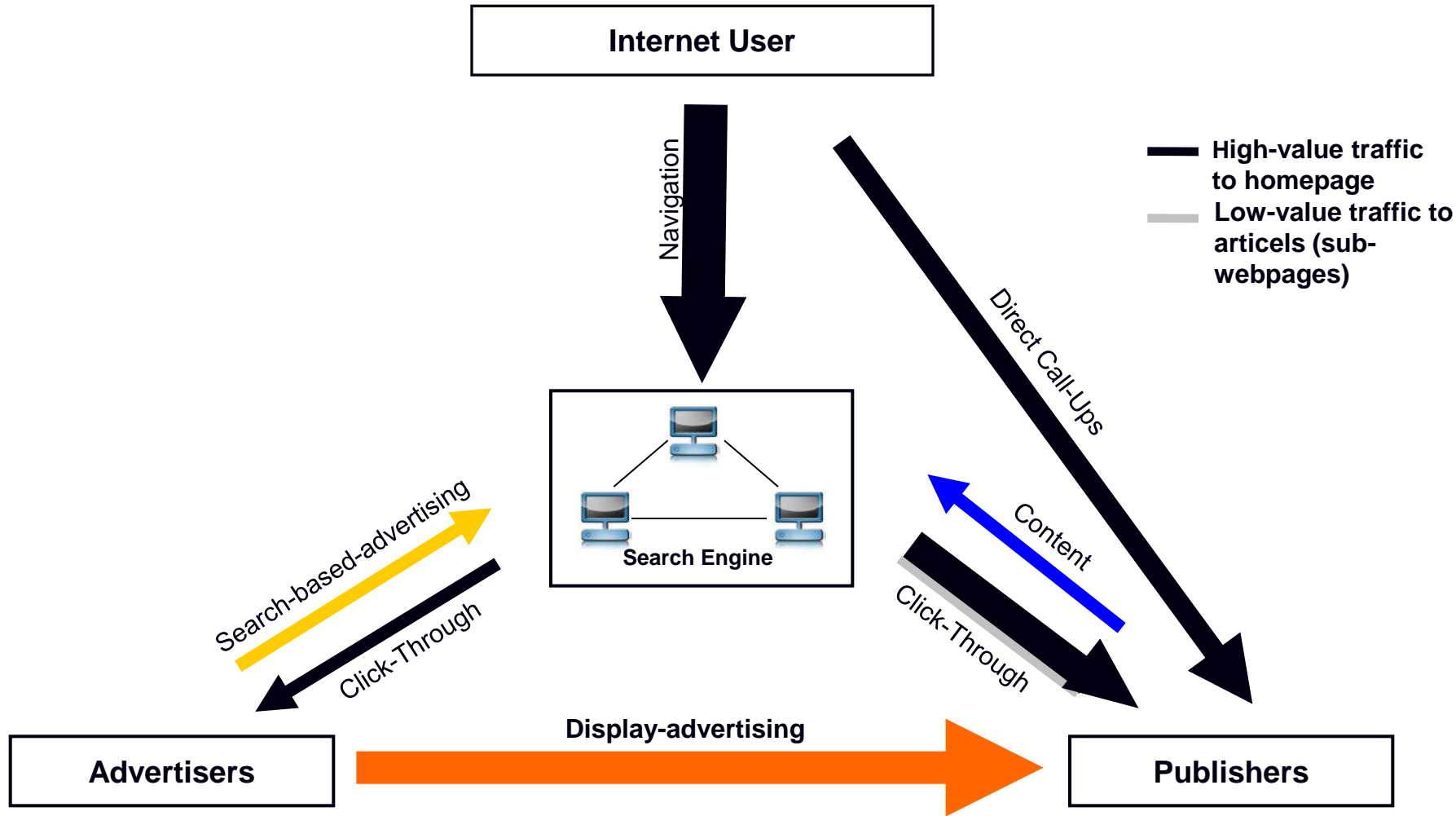


Press Publishers and Search Engines – Between Cooperation and Confrontation

Machine Readable Rights and the News Industry
IPTC Workshop, March 12, 2013
Dr Thomas Höppner, Berlin



Original business model of search engines – neutral navigation



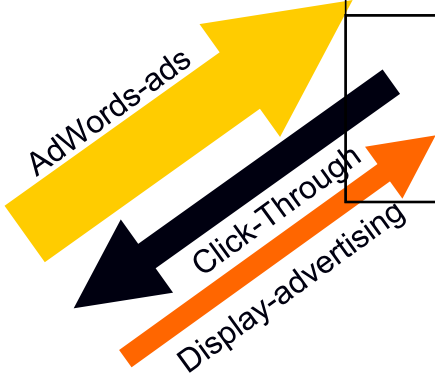
„User filter“ of search engines - example of Google

Internet User

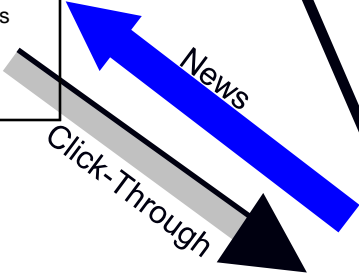
OLSWANG

- High-value traffic to homepage**
- Low-value traffic to articles (sub-pages)**

Google Web Search		Google News	Google Reader	Google Alert
classic homepage	iGoogle			
<p>News (Link)</p> <p>→</p> <p>News for XYZ (Universal)</p>	<p>News (Link)</p> <p>→</p> <p>Google News Gadget</p> <p>News for XYZ (Universal)</p>	<p>News-Snippets</p> <p>← + →</p> <p>hosted news and photos from press agencies</p>	<p>News (Link)</p> <p>←</p>	<p>News (Link)</p> <p>←</p>



- Diversion of users
- Delivery of news



Advertisers



Press Portals

Traffic lost through news-aggregation

OLSWANG

„We did a study of traffic on several sites that aggregate purely a menu of news stories. **In all cases, there was at least twice as much traffic on the home page as there were clicks going to the stories that were on it.** In other words, a very large share of the people who were visiting the site were merely browsing to read headlines“

(Mishkin, The Fallacy of The Link Economy)

„**Online readers read shallow but wide**, while at the same time pursuing selected topics in depth. **More than three times as many briefs as articles were called up.** Sessions averaged 34 minutes.“

(Standord-Poytner Project 2001)

“**A full 44 percent of visitors to Google News scan headlines without accessing newspaper’s individual sites**” - “Nearly Half of News Users Bypass Newspaper Sites in Favour of Google”

(Outsell Study 2009)

Alternatives for publishers

- **Opting-out using the Robots Exclusion Protocol (REP) – Brazil / Italy / US (FTC)**

- Using the robots.txt and Robot-Meta-Tags to block Google
- Separate blocking of Google News and Google Web Search

- **Ancillary copyright law for press publishers – Germany**

- Protection of publishers' content against any copying in whole or in part (including few words) by search engines
- Protection independent of violation of author's copyrights

- **„Link-Tax“ – charge for links to press articles – France?**

- Charge for every link to a news article
- Publishers „forced“ to guarantee free access to their content

- **Negotiating / cooperating with search engines – Belgium and France?**

- On technical solutions, eg more opt-out options or rights expression languages (ACAP, RightsML)
- On financial participation, eg sharing of news-related advertising revenues

Robots Exclusion Protocol

Why it does not suffice I

You cannot block..	..without also blocking
Google News	„News to...“ Universals in Web Search
Google News	Google Alert
Snippets	(Instant) Previews
Google Web Search	Affiliates using the Google Index (eg AOL, Gmx.com, MySpace, T-Online.de)
Google form crawling one item (eg an article) on a webpage	all other items on that webpage

Robots Exclusion Protocol

Why it does not suffice II

- **Robots.txt – index or no-index**

- Only option: disallow indexing of entire site, directory, webpage, specific image from Google Images, files of a specific file type
- Other actions (display, archive etc) can only be excluded via Meta Tags

- **Robots Meta Tags – entire webpage or nothing**

- *“When we see the noindex meta tag on a page, Google will completely drop the page from our search results”. (Google Webmaster-Tools)*
- *“Can the robots meta tag be used to block a part of a page from being indexed?: No, the robots meta tag is a page-level setting”. (ibid)*

- **No communication of usage conditions**

- Neither the robots.txt nor Robots Meta Tags can be used to make usage dependent on certain terms and conditions

Robots Exclusion Protocol

Why it does not suffice III

- **Google's Webmaster-Tools: access to content is crucial for ranking**

- *"Provide high-quality content on your pages, especially your homepage. This is the single most important thing to do."*
- *"Make your site easily accessible"*
- *"Blocking Google from crawling a page is likely to decrease that page's ranking or cause it to drop out altogether over time. It may also reduce the amount of detail provided to users in the text below the search result."*

- **SEO-Expert (webconfs.com/15-minute.seo.php)**

- *"If indexing of a considerable portion of the site is banned, this is likely to affect the non-banned part as well because spiders will come less frequently"*

Minimum Requirements for a Rights Management System

- **Acceptance of machine-readable language** for communication of standard terms & conditions for usage of contents
- Rights communication **for individual works** not web pages
- **Usage conditions** that take publishers' interests into account, eg:
 - allow indexing but specify a time limit for storing of content;
 - allow free use for some purpose but require a licence fee for others;
 - express constraints upon the presentation of content to end-users, eg the maximum length of snippets
- **No direct or indirect punishment** for using rights management system
- **No preferential treatment** of content-providers for waiving IP-rights



For more information
please contact:

Dr Thomas Höppner

+49 30 700171 176

thomas.hoppner@olswang.com