



UvA-DARE (Digital Academic Repository)

Search and exploration of X-Rated information (SEXI 2013)

Murdock, V.; Clarke, C.L.A.; Kamps, J.; Karlgren, J.

Published in:
WSDM 2013

DOI:
[10.1145/2433396.2433507](https://doi.org/10.1145/2433396.2433507)

[Link to publication](#)

Citation for published version (APA):

Murdock, V., Clarke, C. L. A., Kamps, J., & Karlgren, J. (2013). Search and exploration of X-Rated information (SEXI 2013). In *WSDM 2013: proceedings of the 6th ACM International Conference on Web Search and Data Mining: February 4-8, 2013, Rome, Italy* (pp. 795-796). New York: ACM.
<https://doi.org/10.1145/2433396.2433507>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Search and Exploration of X-Rated Information (SEXI 2013)

Vanessa Murdock
vanessa.murdock@yahoo.com

Charles L.A. Clarke
University of Waterloo
claclarke@plg.waterloo.edu

Jaap Kamps
University of Amsterdam
kamps@uva.nl

Jussi Karlgren
Gavagai Stockholm
jussi@gavagai.se

ABSTRACT

Adult content is pervasive on the Web, has been a driving factor in the adoption of the Internet medium. It is responsible for a significant fraction of traffic and revenues, yet rarely attracts attention in research. We propose that the research questions surrounding adult content access behaviors are unique, and we believe interesting and valuable research in this area can be done ethically. The workshop on *Search and Exploration of X-Rated Information* (SEXI) addresses these issues for information access tasks related to adult content.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Clustering, Information filtering, Search process

General Terms

Algorithms, Measurement, Experimentation, Human Factors

Keywords

Adult content, user generated content

1. INTRODUCTION

Adult content is pervasive on the Web, has been a driving factor in the adoption of the Internet medium. It is responsible for a significant fraction of traffic and revenues, yet rarely attracts attention in research [1,5]. The scientific community has spent considerable energy studying user-generated content and information access on the Web, to the exclusion of adult content. This is understandable, as the topic is distasteful to some, and requires special legal and ethical considerations when asking employees, contractors and students to analyze and process the data. Furthermore, the methods that work for other types of information access behavior are assumed to work for all types of content, including adult content.

We propose that this is an incorrect assumption. In fact, even core concepts such as relevance and diversity, which are fundamental to any application involving information seeking and access, are defined differently for adult content. Adult queries frequently fall outside of the taxonomy of queries (informational, transactional, navigational) that applies to standard Web queries. Users searching for adult content frequently have an entertainment need, rather than an information need. Thus, because of the nature of the content, the user may be more satisfied with multiple similar images, than with a set of search results that capture different meanings of the query terms. Furthermore, understanding that a user is searching for a term in an adult context often disambiguates the term.

For example, consider the query "bikini". In a non-adult context, the user might be querying about a ham and cheese sandwich, or may be interested in viewing catalog photos of bikinis to purchase, or perhaps would like images of the Bikini Islands. Since the search engine cannot be sure, one strategy is to represent multiple senses of the query in the results presented to the user. In an adult context, catalog photos, or images of sandwiches and islands will be more than an annoyance. Images of people wearing bikinis, although repetitive and not representing a diverse sense of the query term, is what the user is expecting.

Equal in importance to serving adult content in the best possible way, is the issue of avoiding serving adult content to those who are not looking for it. Many innocuous terms (such as "snake," "cougar," "swimsuit") have adult connotations. Understanding when a person is uploading adult content is important. Often the only information available to determine the adultness of an image or video are the vague tags the user applies. Complicating the interpretation of the tags is that adult content may be described euphemistically with ordinary nouns that reflect a particular visual imagery. Similarly, when a person issues a query, it is not always clear whether they are searching for adult content, and it is extremely important for the search engine to understand this before serving adult content to a person who is not expecting it.

2. OPEN QUESTIONS

We seek a greater understanding of the particular issues in accessing adult content, especially user-generated adult content on the Web. The focus of the workshop will be to put this area of research on the agenda, and explore the basic research questions that should be addressed in the field, the types of data needed for research, and the barriers to doing research in this area. Due to the lack of attention to this area of research there are many open questions.

Classification Even researchers and search applications not interested in adult content will have to deal with it in order to avoid it--presenting adult content to innocuous searchers is clearly a massive failure both for the individual searcher as well as for the reputation of the service. What are automatic methods for identifying adult content, in particular adult user-generated content? How can we identify adult content in video, images, and text? What is the best way to identify adult query intent, and deal with ambiguous requests? What are the appropriate ad placement strategies in adult content?

Access Access to adult content seems to require a different approach than the ubiquitous navigation search---with searchers exhibiting an exploratory information seeking behavior, characterized by a diverse set of relevance criteria. How should adult content be ranked? How should search, exploration, and recommendation be balanced? How does searching adult content relate to search on adult chat sites and social networks? Is there a benefit to personalizing adult content?

Evaluation Given the distinct nature of adult content and the diverse relevance criteria, appropriate evaluation is crucial. What is a relevant result, and what are suitable metrics for relevance? Is adult content a recall-oriented, or precision-oriented task? What is the right level of evaluation---individual requests or whole search sessions? What is similarity and diversity in adult content? How important is the avoidance of failure, relative to success? Are searchers for adult content more tolerant of non-relevant results?

Ethics What are the ethical issues in working with adult content in an academic environment? What are the ethical implications for the search industry, given that it partly facilitates the online adult industry? How can adult material be made available so as to promote responsible behavior through the whole chain from production to consumption? Is adult user-generated content more ethical than professionally produced media?

The outcome will be to define a set of research areas, to elucidate the special issues surrounding the access of user-generated adult content. We discuss a set of best practices for working with this data in an academic environment, and propose a research agenda for the near future.

3. RELEVANCE

The adult entertainment industry is in decline. Because of an increasing amount of amateur adult content that is uploaded to the Web every day, it is difficult for the professional industry to remain competitive [2,3,4,6]. Although the industry is in decline, the demand for online porn, in particular free online porn, remains constant. This transition moves adult content to the center of the search industry.

Searches for adult content represent a significant portion of search engine query logs. The industry generates billions of dollars in revenue, from paid sites as well as contextual advertising. On the other side of the coin, porn sites have a high volume of traffic, but are also one of the primary mechanisms for spreading malware. For these reasons, the workshop is relevant and important to improve information access for every user, and not just users searching for porn.

As stated in the Introduction, user-generated adult content is quickly becoming the predominant adult content on the web. Although it lacks high quality production values, it is sought by users because it is frequently free, and represents a reality and a diversity of human behavior that may be lacking in professionally produced content. According to an article in the Los Angeles Times [4]: "Sites like Pornhub, YouPorn and RedTube attract more users than TMZ and the Huffington Post. The porn sites are even bigger than Pirate Bay, the top portal for illegal downloads of movies, TV shows and music...." As these are sites that allow regular users to upload their own adult content, this represents a paradigm shift in the porn industry. In fact, it is putting the professional porn industry out of business. From the same article, "Reliable revenue and employment figures for the adult industry don't exist, since no analysts or economists track it. Adult Video News estimated in 2006 that it was worth \$13 billion, but Paul Fishbein, editor of the trade publication, said the number was 'an educated guess.'"

In short, this topic is central to the Web in all aspects, ranging from Web search to data mining and classification of web data. It is the topic that everyone knows about, but no-one speaks about. This workshop is the first of its kind in the information retrieval research community.

4. REFERENCES

- [1] Azzopardi, L. Searching for unlawful carnal knowledge. In Belkin, N.J. et al. (eds.). *Proceedings of the SIGIR'11 Workshop on "entertain me": Supporting Complex Search Tasks*, pp 17-18. ACM Press, 2011.
- [2] Economist. Hard times: A big industry in northern los angeles is among the worst hit by the recession. *The Economist*, September 10, 2009. <http://www.economist.com/node/14416740> .
- [3] Friess, S. Porn industry seats recession, piracy. *AOL News*. January 9, 2011. <http://www.aolnews.com/2011/01/09/porn-industry-facing-hard-times-in-struggling-economy>
- [4] Fritz, B. Tough times in the porn industry. *The Los Angeles Times*. August 10, 2009. <http://www.latimes.com/news/local/la-fi-ct-porn10-2009aug10,0,3867866,full.story> .
- [5] Halavais, A. C. Small pornographies. *SIGGROUP Bulletin*, 25(2):19-22, 2005. <http://doi.acm.org/10.1145/1067721> .
- [6] Theroux, L. Louis Theroux on porn: The decline of an industry. *BBC News Magazine*, June 8, 2012. <http://www.bbc.co.uk/news/magazine-18352421>.