Chapter 5
Creating and Analysing a Social Network Built from Clips of Online News

Álvaro Figueira
CRACS & INESC TEC, Universidade do Porto, Portugal

José Devezas
CRACS & INESC TEC, Universidade do Porto, Portugal

Nuno Cravino
CRACS & INESC TEC, Universidade do Porto, Portugal

Luis-Francisco Revilla
School of Information, University of Texas at Austin, USA

ABSTRACT
Current online news media are increasingly depending on the participation of readers in their websites while readers increasingly use more sophisticated technology to access online news. In this context, the authors present the Breadcrumbs system and project that aims to provide news readers with tools to collect online news, to create a personal digital library (PDL) of clips taken from news, and to navigate not only on the own PDL, but also on external PDLs that relate to the first one. In this article, the authors present and describe the system and its paradigm for accessing news. We complement the description with the results from several tests which confirm the validity of our approach for clustering of news and for analysing the gathered data.

INTRODUCTION
News media are at a point of historic transition. The conjuncture of digital and social media is merging the roles of readers and providers. Readers are increasingly participating in the news media cycle, where news are written, published, commented, associated with other news, improved, and published again. More than ever, the future of news depends on harnessing the participation of readers in the global process of production and consumption of news.

DOI: 10.4018/978-1-4666-4062-7.ch005
Many new sites and bloggers try to bind their readers to the news sites using different approaches and means. Usually, the comments are the most participated but, the “like” or the “+1” used in social networks are also increasing their spread among news media. Other less, participated means are the connections to and from blogs, or even the editing of short stories as in Scoop (www.scoop.it). Given this quantity of links, connections and information provided by readers, it would be losing an opportunity if news producers wouldn’t take advance of this information in order to better understand reader’s needs, interests in order for further development of the stories or even improvements.

In this article we present and describe the Breadcrumbs system whose goal is to capitalize on the participation of the general public in the production of news by creating bridges between online news and the “Social Web” while stepping over the traditional techniques of natural language processing and its associated complexity. The project builds on the use of Social Web tools that we created for gathering the opinions of readers for the news that are interesting to them, and then creating a semantically organized model of the readers’ opinions. In particular, Breadcrumbs focuses on:

1. Collecting news fragments from the Web.
2. Semi-automatically organizing those fragments.
3. Aggregating the fragments across readers and building the social network of news.
4. Anonymously inferring relationships between readers.
5. Inferring relationships between news.

In order to accomplish these tasks, we use various inference and interaction approaches. We combine automatic and user-mediated approaches to yield better results than either approach in isolation – our rationale is: automatic mechanisms can handle extremely large amounts of data, and people can provide insights which are difficult to identify with automatic mechanisms.

The remaining of this article is structured as follows: in this next section we describe our insight in creating the Breadcrumbs system and try to ground our beliefs in what concerns the viability and opportunity of a social network formed by news clips. In section 1 we detail some of the most important concepts to prepare the reader to understand the state of the art in the area. Section 2 is devoted to the full description of our system, which is then analysed through several tests described in section 3, along with our findings. In section 4 we present our conclusions.

The Breadcrumbs Paradigm

As evidenced by the success of social bookmarking systems (e.g., delicious.com), people like to keep track of digital information items, storing and collecting them, such that they can be accessed, reviewed, or used later. In Breadcrumbs we extend this by allowing people to keep track of news at a highly fine-grained detail level. Breadcrumbs lets readers select news stories fragments from any news site, blog, etc., collect and store them in their own Personal Digital Library (PDL). The fragments can also be annotated with tags and comments. While we feel that currently people are more willing to use tags for better content discovery later on, the comments can provide some sort of metadata that will help the user understand why the fragment is important or why it was actually collected. Therefore, this information is used by the user to create his/her own organization of the clips, or to provide the clips with some kind of mental model for his/her views of the PDL. Concurrently with this, Breadcrumbs also used the tags and comments to learn new characteristics of the collected clips in order to enhance the automatically clustering of the PDL and to infer relations at the system-wide PDL level.

Actually, while each PDL represents the individual perspective of a reader, we believe that by