Article 25fa pilot End User Agreement

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) ‘Article 25fa implementation’ pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication other than authorised under this licence or copyright law is prohibited.

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 contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library
Radboud University
Copyright Information Point
PO Box 9100
6500 HA Nijmegen

You will be contacted as soon as possible.
Hypertext as a Tool for Exploring Personal Data on Social Media

Eelco Herder
Radboud Universiteit Nijmegen
Institute for Computing and Information Sciences
Nijmegen, The Netherlands
eelcoherder@acm.org

Daniel Roßner
Hof University
Institute of Information Systems
Hof, Germany
daniel.rossner@iisys.de

Claus Atzenbeck
Hof University
Institute of Information Systems
Hof, Germany
claus.atzenbeck@iisys.de

ABSTRACT
Social networks such as Facebook are required to provide users with their personal data. However, these dumps do not provide users insight in overarching themes in their online behavior. In this poster, we discuss the development of Mother, a spatial hypertext system for visual data exploration. First insights include that the less obvious connections are more interesting and relevant to the user than very close semantic or temporal connections.

CCS CONCEPTS
• Human-centered computing → Hypertext / hypermedia; Graphical user interfaces; User centered design; • Software and its engineering → Software infrastructure.

KEYWORDS
spatial hypertext; social networks; user profiles; GDPR; data exploration

ACM Reference Format:

1 INTRODUCTION
Social networks are widely spread. The Digital 2020 report1 counts 2.5 billion active Facebook users, which makes Facebook the world’s most-used social platform. The vast amount of information communicated by a huge group of people on a single social media platform creates significant advantages to the platform provider [2].

Increasingly, governments introduce regulations that protect citizens against the company’s market power or lobbyists. The European General Data Protection Regulation (GDPR) [4] opens, among others, the right to EU citizens to request their data from platform providers.

In this poster, we present our system Mother [1] as a hypertext tool for exploring social media data. The goal is to increase the users’ awareness of what data and, more importantly, which associations between information units the platform provider stores. We will use Facebook data dumps as an example for social media platforms.

2 BACKGROUND
Personal data is used by platforms for a variety of purposes, varying from personalization and recommendation to monetization, for example via advertisements and nudges to continue visiting the platform [5].

The introduction of the General Data Protection Regulation, the GDPR [4], in Europe provides end-users with means for requesting transparency. Dumps of personal data or posts from social media platforms may be voluminous. Therefore, they generally need to be teased apart into smaller coherent informational units and associated in a semantically meaningful way.

From its very beginning [3], hypertext has been considered a medium for representing human associations, a medium for humans to express their (interconnected) ideas. A special type is spatial hypertext [6], which follows a desk-on-table metaphor: similar to physical paper notes on a desk, objects can be moved on a 2D canvas.

The big advantage of using spatial hypertext is that structures can be created and modified at ease and at a low cognitive load [7]. As such, spatial hypertext helps in exploring unknown knowledge by supporting creating, modifying, or destroying contexts of information to which the system reacts to.

3 SYSTEM DESIGN AND DEVELOPMENT
Following the GDPR regulations, Facebook allows its users to obtain a copy of their personal data as a simple download2. The user receives an archive with files that contain, among others, the user’s own posts, comments, likes and reactions to posts by others (both friends and Facebook pages), search history, lists of friends, subscribed groups and pages, and interaction with advertisements.

The textual contents of the posts as well as post frequency statistics can provide rich material for users to obtain insight in and to reflect on their Facebook usage, including the reporting of life events, work-related announcements, discussions with friends, shared silly pictures or memes, and interaction with advertisements.

As a first step towards a visualization of Facebook data, we created a script to process the JSON Facebook post data of a user into a graph-based format, with posts, keywords, months and years as vertices, connected by edges with various weights. The keywords are extracted from the Facebook posts, converted into lowercase, lemmatized, stopwords removed and only keywords that appear in

---

1See https://datareportal.com/reports/digital-2020-global-digital-overview, data updated to 25 January 2020

2https://www.facebook.com/settings?tab=your_facebook_information
at least 5 posts are stored, in order to keep the number within limits. Edges between posts, years, months and keyword were created and weighted based on tf-idf and/or co-occurrence.

From a user’s point of view, the principle of Mother is that one selects a single entity (in our case, a Facebook post), which is then displayed along with related entities (other posts and keywords) as recommendations that can be followed in order to create a narrative. An example is shown in Fig. 1.

The first developed application area of Mother concerned the movie domain, where movies are connected with one another through actors, genres and other entities [1]. This content-based approach made sense in the movie domain, but applied to Facebook posts they resulted in relations that were too obvious for the user (e.g., birthday wishes were related to other birthday wishes). After all, there is a difference between exploring an unknown domain, and introspection, the examination or observation of one’s own mental and emotional processes.

For this reason, we decided to only recommend posts based on a combination of content-based and temporal similarity (hour of day, day of week, month), manually tuned and evaluated by the authors in several sessions. Furthermore, we also added the keywords as vertices and related them to the posts. Given the large differences in posting behavior, even between the authors, it was concluded that no optimal a-priori values could be found. Instead, we opted for a configuration that led to the first author’s observation that he typically posted his reflections on a day in the early evening; in addition, some randomness was added in order to prevent users to get locked in a small number of favorite themes.

4 CONCLUSIONS AND FUTURE WORK
An important lesson learned during the design of the system and exploration of the author’s own profiles is that there is a difference between exploring an unknown domain (e.g., movies) and introspective exploration of one’s own activities and posts: for unknown domains, close semantic relations (such as actor X plays in movie Y) are meaningful and useful, but when exploring one’s own activities, these relations turn out to be to obvious to be useful.

As future work, we aim to improve the interface, so that the posts are presented along with photos and links, if present. Furthermore, we will use similar approaches for exploring different parts of a user’s Facebook profile: we expect that users’ remembrance of their own comments to friends’ posts and page contents is far lower than of their own posts and therefore can provide more insight in which triggers one is sensitive to.

REFERENCES