Namelings
Discover Given Name Relatedness Based on Data from the Social Web

Folke Mitzlaff[1] and Gerd Stumme[2]

[1] mitzlaff@cs.uni-kassel.de
[2] stumme@cs.uni-kassel.de
Knowledge and Data Engineering Group (KDE), University of Kassel
Wilhelmshöher Allee 73, D-34121 Kassel, Germany

Abstract. During the exhausting search for a given name for the yet unborn baby, the idea of a name recommendation system based on relations mined from the "*social web*" was born. This demonstration paper presents the Nameling[3], a recommendation system, search engine and academic research platform for given names, which attracted more than 30,000 users within four months, underpinning the relevance of the task and associated research questions.

Keywords: Given Names, Network Analysis, Recommendation System

1 Introduction

Whoever had to chose a given name, knows how challenging it is to find a suitable name which fits to the personal preference as well as the social environment. There is a huge bibliography on books, listing given names in alphabetical order, as well as dozens of respective websites. But finding a suitable name in such a list is an exhausting task. Out of several thousand names, only a small fraction of names is typically relevant to the reader.

In different contexts, recommendation systems are subject to scientific research, as, e. g., finding relevant annotations [5], recommending products [3] or suitable movies [2]. So far, the task of finding relevant given names is not formally investigated, though relevant in practice. Based on data observations from the social web, the Nameling generates such recommendations, enabling users to search for suitable names and browse through a list of more than 35,000 given names.

2 Basic Concepts

The Nameling is designed as a search engine and recommendation system for given names. The basic principle is simple: The user enters a given name and gets a browsable list of "relevant" names, called "*namelings*". Figure 1a exemplarily shows namelings for the classical masculine German given name "Oskar".

The list of namelings in this example ("Rudolf", "Hermann", "Egon", . . .) exclusively contains classical German masculine given names as well. Whenever an according article in Wikipedia exists, categories for the respective given name are displayed,
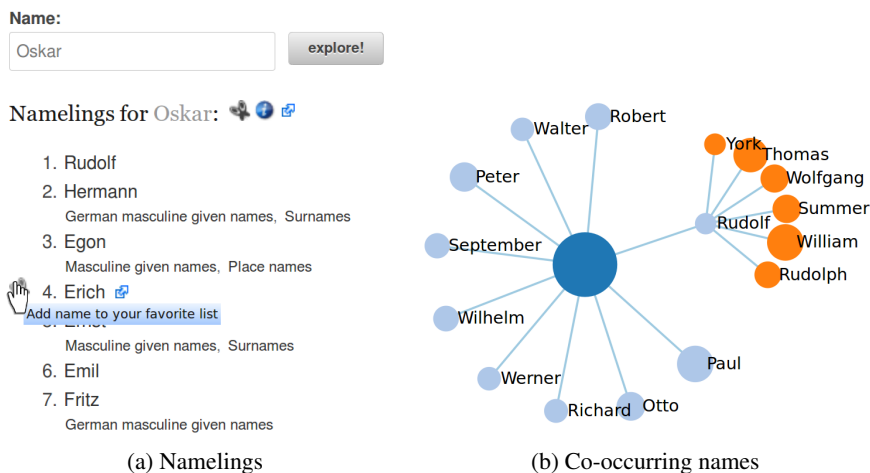
---

[3] http://nameling.net

(a) Namelings

(b) Co-occurring names

Fig. 1: The user queries for the classical German given name "Oskar".

as, e. g., "*Masculine given names*" and "*Place names*" for the given name "Egon". Via hyperlinks, the user can browse for namelings of each listed name or get a list of all names linked to a certain category in Wikipedia. Further background information for the query name is summarized in a corresponding details view, where, among others, popularity of the name in different language editions of Wikipedia as well as in twitter is shown. As depicted in Fig. 1b, the user may also explore the "neighborhood" of a given name, i. e., names which co-occur often with the query name.

From a user's perspective, the Nameling is a tool for finding a suitable given name. Accordingly, names can easily be added to a personal list of favorite names. The list of favorite names is shown on every page in the Nameling and can be shared with a friend, for collaboratively finding a given name.

3   Background

With the rise of the so called "Web 2.0", various social applications for different domains emerged, offering a huge source of information and giving insight into social interaction and personal attitudes. The basic idea behind the Nameling was to discover relations among given names, based on such user generated data. In this section, we briefly summarize how data is collected and how relations among given names are established.

The Nameling is based on a comprehensive list of given names, which was initially manually collected, but then populated by user suggestions. It currently covers more then 35,000 names from a broad range of cultural contexts. For different use cases, three different data sources are respectively used, as depicted in Fig. 2:

Wikipedia: As basis for discovering relations among given names, a co-occurrence graph is generated for each language edition of Wikipedia separately. That is, for each

WIKIPEDIA   twitter   facebook.

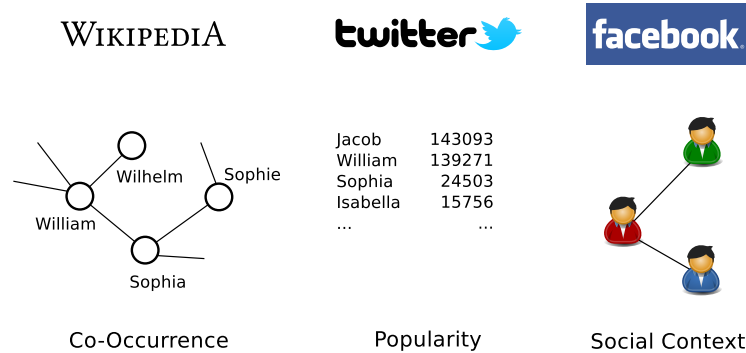Co-Occurrence   Popularity   Social Context

Fig. 2: The Nameling determines similarities among given names based on co-occurrence networks from Wikipedia, popularity of given names via twitter and social context of the querying user via facebook.

language, a corresponding data set is downloaded from the Wikimedia Foundation[4]. Afterwards, for any pair of given names, the number of sentences where they jointly occur is determined. Thus, for every language an undirected graph is obtained, where two names are adjacent, if they occur together at least in one sentence within any of the articles and the edge's weight is given by the number of such sentences.

Relations among given names are established by calculating a vertex similarity score between the corresponding nodes in the co-occurrence graph. Currently, namelings are calculated based on the cosine similarity, which performed best in according experimental evaluations [4].

twitter: For assessing up-to-date popularity of given names, a random sample of tweets in twitter[5] is constantly processed via the twitter streaming api[6]. For each name, the number of tweets mentioning it is counted.

facebook: Optionally a user may connect the Nameling with facebook[7]. If the user allows the Nameling to access his or her profile information, the given names of all contacts in facebook are collected anonymously. Thus, a "social context" for the user's given name is recorded. Currently, the social context graph too small for implementing features based on it, but it will be a valuable source for discovering and evaluating relations among given names.

4 Emerging Usage Data & Research Questions

Beside being a tool for parents-to be, the Nameling also serves a research platform. The choice of a given name is influenced by many factors, ranging from cultural background

---

[4] `http://dumps.wikimedia.org/backup-index.html`
[5] `http://twitter.com`
[6] `https://dev.twitter.com/docs/api/1/get/statuses/sample`
[7] `http://www.facebook.com`

over social environment to personal preference. Accordingly, the task of recommending given names is per se subject to interdisciplinary considerations.

Within the Nameling, users are anonymously identified via a cookie, that is, a small identification fragment which uniquely identifies a user's web browser. From now on we talk about users relative to this identification, ignoring the fact that users may use different browsers and/or computers.

Around 37,000 users issued more than 330,000 search queries within the time range of consideration (2012-03-06 until 2012-07-23). For every user, the Nameling tracks the search history, favorite names and geographical location based on the user's ip address and the GeoIP[8] database. All these footprints together constitute a multi-mode network with multiple edge types. Analyzing this graph (or one of its projections) can reveal communities of users with similar search characteristics or cohesive groups of names, among others.

Most importantly, recommendation systems can be personalized based on the Nameling's usage data via association rule mining [1] or collaborative filtering [6]. But also new approaches can be applied and evaluated, e. g., by considering a users geographical location. Furthermore, the usage data can also be used as a reference for evaluating and improving the process of discovering name relatedness.

5   Conclusion

This demonstration paper introduced the Nameling, a search engine and research platform for given names. From a user's perspective, many more features are desirable - but beforehand, methods for mining relatedness of given names must be evaluated and specialized recommendation systems developed. The analysis of given names and associated social background information is predestined for interdisciplinary considerations, whereby the usage data which accrues at the Nameling may serve as a valuable source of reference data.

References

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th international conference on Very Large Data Bases (VLDB'94)*, pages 478–499. Morgan Kaufmann, September 1994.
2. J. Golbeck and J. Hendler. Filmtrust: Movie recommendations using trust in web-based social networks. In *Proceedings of the IEEE Consumer communications and networking conference*, volume 96. Citeseer, 2006.
3. G. Linden, B. Smith, and J. York. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.
4. F. Mitzlaff and G. Stumme. Mining relatedness among given names based on social co-occurrences. Technical report, 2012. Submitted for Publication.
5. J. J. Pazos-Arias, A. Fernández Vilas, and R. P. Díaz Redondo. Recommender systems for the social web, 2012.
6. L. Terveen and W. Hill. Beyond recommender systems: Helping people help each other, 2001.

---

[8] `http://www.maxmind.com/app/ip-location`