Tags, Titles or Q&As? Choosing Content Descriptors for Visual Recommender Systems

Belgin Mutlu
Know-Center GmbH
bmutlu@know-center.at

Eduardo Veas
Know-Center GmbH
eveas@know-center.at

Christoph Trattner
MODUL University Vienna
christoph.trattner@modul.ac.at

ABSTRACT

In today's digital age with an increasing number of websites, social/learning platforms, and different computer-mediated communication systems, finding valuable information is a challenging and tedious task, regardless from which discipline a person is. However, visualizations have shown to be effective in dealing with huge datasets: because they are grounded on visual cognition, people understand them and can naturally perform visual operations such as clustering, filtering and comparing quantities. But, creating appropriate visual representations of data is also challenging: it requires domain knowledge, understanding of the data, and knowledge about task and user preferences. To tackle this issue, we have developed a recommender system that generates visualizations based on (i) a set of visual cognition rules/guidelines, and (ii) filters a subset considering user preferences. A user places interests on several aspects of a visualization, the task or problem it helps to solve, the operations it permits, or the features of the dataset it represents. This paper concentrates on characterizing user preferences, in particular: i) the sources of information used to describe the visualizations, the content descriptors respectively, and ii) the methods to produce the most suitable recommendations thereby. We consider three sources corresponding to different aspects of interest: a title that describes the chart, a question that can be answered with the chart (and the answer), and a collection of tags describing features of the chart. We investigate user-provided input based on these sources collected with a crowd-sourced study. Firstly, information-theoretic measures are applied to each source to determine the efficiency of the input in describing user preferences and visualization contents (user and item models). Secondly, the practicability of each input is evaluated with content-based recommender system. The overall methodology and results contribute methods for design and analysis of visual recommender systems. The findings in this paper highlight the inputs which can (i) effectively encode the content of the visualizations and user's visual preferences/interest, and (ii) are more valuable for recommending personalized visualizations.

CCS CONCEPTS

•**Information systems → Recommender systems;**

HT '17, Prague, Czech Republic

DOI: http://dx.doi.org/10.1145/3078714.3078741

KEYWORDS

Recommending visualizations; information theory; user modeling; personalization

ACM Reference format:
Belgin Mutlu, Eduardo Veas, and Christoph Trattner. 2017. Tags, Titles or Q&As? Choosing Content Descriptors for Visual Recommender Systems. In *Proceedings of HT '17, Prague, Czech Republic, July 04-07, 2017,* 10 pages.
DOI: http://dx.doi.org/10.1145/3078714.3078741

1 INTRODUCTION

The effectiveness of personalized recommender systems (RS) highly depends on user and item profile completeness and accuracy [17]. Hence, regardless of the approach, the key factor in personalized recommendations is the decision about how to exploit the relevant information about the user and items, and more important, which information better describes the properties of items and the preferences/needs of a user. Collaborative filtering approaches for instance use ratings, while content based approaches build on user-provided input, typically in form of tags and comments. Rating is fast, simple, and effective in communicating user preferences also in the context of visualizations [16], but it does not indicate much about goals or intentions of the user regarding the item. Annotating visualizations with tags brings extra benefits, as a user indicates her insights and interpretation of the data being visualized, i.e., issuing details with keywords pulled from a personal vocabulary [15, 26, 27]. Hereby, visualizations are organized for later retrieval.

There are two caveats in these approaches to personalizing visualizations: 1) people are often reluctant to give a feedback, 2) ratings and tags forego information about the context where the item was used. Unless the benefit is evident, users rarely engage in tagging or rating items. This is true in the context [4, 26] and can be more acute for visualizations where the user is possibly engaged in a thought process that would be interrupted by rating/tagging. More importantly, a single rating does not tell much about goals or intentions of the user. Whereas tags encode features of the item, it is not evident that users will include their task or intentions when tagging a chart. Our working question is: can we use alternative sources to derive item descriptions suitable for recommendation?

In the context of visualizations, user's provided input (annotations) can take other forms. For instance, it is common for user to pose a question that is answered with a visualization, or to define a title and description for the visualization in form of a caption. We consider these two alternative sources of information (titles, questions&answers (Q&As)) as potential descriptors both of item and user intentions. To investigate how effectively they encode information, each information source (tags, titles, Q&As) is characterized using information-theoretic measures, such as entropy,

conditional entropy and mutual information, as suggested in Chi et al. [3, 11]. Finally, using each of these sources separately, we build models for user and item profiles to recommend personalized visualizations applying a content based recommender. This allows us to obtain insights and draw general conclusions about the drawbacks and benefits for each source as input for the visual recommender systems. The input data for our studies was obtained with a crowd-sourced experiment involving 47 participants that had to provide accurate description of each visualization in forms of tags, a title and a question it may answer.

In a nutshell this paper makes the following contributions. We propose a framework to assess the encoding power of different textual information sources in describing user preferences and visualizations. The framework is used in a thorough analysis of different kinds of user-provided input characterizing data models for user and visualizations. We derive insights on how their nature impacts the generation of personalized visualization recommendations.

2 RELATED WORK

One of the key concerns in personalizing recommendations is building personalized profiles of individual users and candidate items. These profiles constitute models of (i) individual user characteristics describing what the user needs and prefers— user model, and (ii) item characteristics describing what the items represent, their content respectively— item model. Yet content-based recommender systems try to define personalized recommendations by matching up the attributes of the user model with the attributes of the item model. However the following questions arise: (i) which source of information is most effective at encoding user preferences and item characteristics, (ii) which source of information yields the more accurate recommendations, and finally (iii) how to acquire this information from the user.

In traditional content based recommendation approaches, systems collect user preferences by explicitly asking users to share their interest and needs, typically in form of tags. Although partially successful, these approaches often suffer on the missing motivation of the user for annotations [5, 13]. However, recent studies on this topic show that user's motivation to annotate resources increases if this provides a navigational aid to the resources [25]. Ricci et al [22], for instance, present a recommender system to help user with searching for travel products. To define recommendations that are closer to user's needs the system asks user to provide critiques in form of textual feedback when one feature of the recommended product is not satisfactory or very important. The authors prove the effectiveness of their system with an empirical study. This also applies for visualizations. When user annotates visualizations, she provides her insights and her interpretation on the data being visualized. Hence, the annotations serve as analysis finding records and personal reminder for later data discovering and analysis tasks [4].

The process of annotating can be considered as an encoding process where the annotations encode the information (facts, features etc.) about the items [25]. However, it depends on the encoding quality of the used annotation type (tags, titles, Q&As) how good a recommender performs. Chi et al. [3] use information-theoretic measures (entropy, conditional entropy, mutual information) to evaluate the encoding power of tags collected from the social tagging site del.icio.us. Using these measures Chi et al. quantify the diversity in tags and documents and the amount of shared information between them. The obtained results finally provide insight into how effective the tags are at encoding documents. Strohmaier et al. [25] use conditional entropy and orphan ratio for measuring and detecting the tacit nature of tagging motivation by analyzing the tag sets produces by 8 different tagging systems regarding to their encoding and descriptive power. The results of their study show that (i) tagging motivation of individuals varies within and across tagging systems, and (ii) user's motivation for tagging has an influence on produced tags and folksonomies. Yi-ling Lin et al. [11] analyze the tags they collected for images in two different tagging conditions (with and without description) on perspectives such as diversity, specificity, quality, similarity and descriptiveness. The analysis mainly covers common text-quality metrics, such as number of unique- and common words of each content and so far.

Regarding to the research question "which source of information defines the more accurate recommendations" the most notable research is provided by Bellogín [1] et al., who try to identify the sources of information (ratings, tags, social contacts, etc.) most valuable for a recommender in a social music service. To do so Bellogín et al. evaluate a number of content based, collaborative filtering and social recommenders on heterogeneous datasets obtained from Last.fm with well-known metrics precision, recall and ranking based matrix. Next, they compare the characteristics of the generated recommendations using non-performance metrics such as coverage, overlap, diversity and novelty between different set of recommendations.

Our paper extends the research on evaluating the power of annotations (tags, ratings etc.) at encoding documents, music tracks and users' interests in these resources into encoding visualizations and users' visualization preferences. Similar to the relevant works we suggest for this purpose information-theoretic measures entropy, conditional entropy and mutual information as these measures have been proposed in many field to assess the diversity of textual content [8, 14, 28]. Furthermore, we address the question of how valuable different source of information are for recommending visualizations by applying a content based filtering approach as this recommendation approach builds on the content features.

3 APPROACH

We developed a recommender system, called *VizRec*, with the purpose to generate personalized visualizations [15, 16]. The schematic overview of the approach is depicted in Figure 1. VizRec responds to a query with a list of personalized visualizations ordered in a top-n sorted manner. The query is a typical free-form text common in search engines (e.g., "most popular movies 2006-2016"). The response to the query is a dataset compiled by a federated system from various associated sources, each with its proprietary data model. Before passing to VizRec, the data are structured after a common data model with a predefined schema. Within VizRec, two recommendation stages take place. First, a rule-based system applies visual encoding guidelines to generate a collection of visualizations appropriate for the data. Second, the collection is sorted

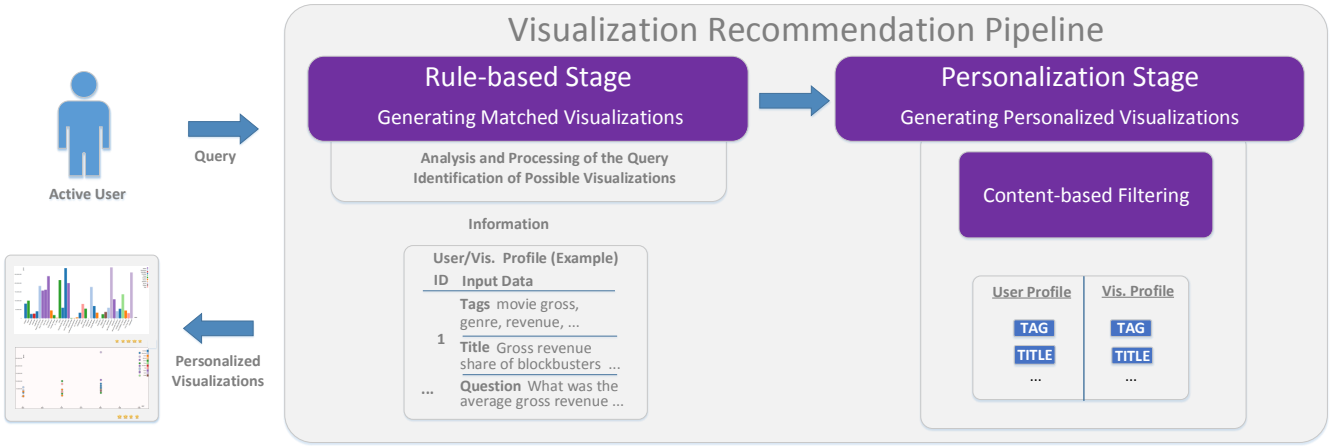

Figure 1: Schematic representation of the visualization recommender: The rule-based stage applies visual encoding guidelines to generate a collection of visualizations appropriate for the data. The personalization stage applies user preferences/profiles (content terms such as tags and titles) and filters the visualizations according to users' needs and interests. This stage also maintains repositories for user preferences/profiles.

and filtered according to user preferences using a content-based recommender system (CB-RS).

Visual encoding guidelines are generic principles that establish relations between visual components of a visualization (e.g., x-axis of a bar chart) and elements of the data (e.g., whether a field is numeric, categorical, or a location, see Section 3.1). A preprocessing unit analyzes the data to structure them in terms of interesting data elements so visual encoding can take place. The three steps to generate personalized visualization recommendations are: (1) preprocessing, (2) visual mapping, and (3) user preference filtering. In the following subsections, we briefly describe each of these units.

3.1 Preprocessing

The preprocessing unit is responsible for extracting and annotating data attributes appropriate for mapping. The input data for VizRec are structured following the specification of the data model described in [18]. The defined specification, concretely, focuses on organizing the different kind of data attributes (content information) extracted from the original sources (ACM digital Library, DBpedia, Mendeley, Europeana etc.). To define the set of appropriate visualizations *VizRec*, first, extracts and analyzes the attributes of the data set being visualized and then categorizes them into standard and/or specific datatypes. The data are categorized into standard datatypes, such as categorical, temporal and numerical – represented by primitive data types string, date and number, respectively. This categorization into primitive datatypes is basically performed by analyzing values of the individual attributes. To do so, the analysis employs a top-down approach, i.e., for a given value it is first decided to which of the aforementioned standard datatypes it belongs. Next, by using gazetteer lists more specialized datatypes are derived, e.g., for spatial information.

Furthermore, the preprocessing unit addresses the task of prior organization of the visualizations into visual patterns each describing one possible combination of visual components of a visualization and data types supported. For instance, two possible patterns for the bar chart are (1) $\{x-axis : string,\ y-axis : number\}$, and (2) $\{x-axis : date,\ y-axis : number\}$. These patterns specify the types of data required for a bar chart to be instantiated. Note that the pattern definition is based on so called Visual Analytics (VA) Vocabulary. For more details about the used vocabulary we refer to our previous paper [15].

3.2 Visual Mapping

The visual mapping process can be considered as a schema matching problem [20]. The basic idea behind schema matching is to figure out a semantic relevance between two objects in schemes under consideration. The result is a mapping comprising a set of elements, each of which indicates that certain elements of schema S1 are related to certain elements of schema S2. In our case, the schemes we deal with are on the one hand the data model which describes the input data, and on the other hand the VA Vocabulary which describes the semantics of the visualizations. Hence, the schema matching in our context produces mappings (possible configurations of a visualization) each of which describes the correspondence between a data attribute of user's current data and a visual component of a visualization. Concretely, the relation from a data attribute to a visual component is valid only if we can establish syntactic correspondences between them. One possibility to identify this is to verify the data type compatibility. The preprocessing unit provides visual patterns for visualizations and the data attributes both including the data types of their elements. Thus, to define a valid mapping the mapping operator simply compares the data types of the visual components and data attributes and builds so the list of plausible mappings. For more details about the mapping algorithm we refer to our previous paper [15].

3.3 User Preference Filtering

To finally filter the generated mapping combinations according to the user's preferences, we employ a content based recommender

system (CB-RS). In a nutshell, the CB-RS generates recommendations by analyzing the relevant content, concretely, the information we know about the active user and the information we extracted from the items (visualizations). The item specific information might be (i) features describing the item, (ii) annotations user applied to the item, or (iii) both features and annotations. Yet, if the visualization is e.g., a bar chart showing the budget per genre (see Figure 2 top left), the features describing this visualization would be the data fields *genre* and *budget* plotted on the $x-axis$-and $y-axis$. Note that these data fields are extracted from the current dataset they therefore not only represent the content of the visualization but also of the dataset.

Following the basic principles of CB-RS, the recommendations are produced based on the content similarity, in our case between the interests of the active user i.e., her profile, and the content information of the candidate items, item profile respectively. An excerpt of these both profiles is given in Figure 1, in the "User/Vis. Profile" block. Generally, a profile is a collection of terms provided to characterize user or items. Thus, for each user in user profile, there is a set of terms describing interests of that particular user. Yet, a user annotates a visualization with terms which describe its content and thus serve as information sources to profile that particular visualization [2, 11]. To take this into account, our recommender defines the item (=visualization) profiles with the aggregated terms supplied by the current user in the past. Note, before we build the profiles we perform a normalization process on the keywords, which involves, (i) removing of commoner morphological and inflectional endings from English words using the Porter stemmer algorithm [9], (ii) removing of stop words (what, how, some, many, etc.) and punctuations (keyword tokenizer), and finally (iii) the lowercase filtering. This step helps to avoid that the words represented in various language forms are interpreted differently [12]. As mentioned, we address in this paper three types of input data models: tags, title and Q&As. So, for each entry in our user profile, we have normalized terms categorized either to tags, titles or question. However, in our item profile we have normalized terms categorized either to tags, titles or answers.

Similarity Estimation and Item Ranking: To determine the correlation between visualizations and users, we transform the content of the user profiles and item profiles into the Vector Space Model (VSM) with the TF-IDF (Term Frequency-Inverse Document Frequency) weighting schema. VSM is a common technique to vectorize the content and in this way to enable analysis tasks, such as classification and clustering for example. In our case, VSM consists of user profile (the current tags, titles or questions) and item profile (tags, titles or answers user applied to the visualizations in the past), both represented in form of vectors. Concretely, using this scheme, each visualization is defined as an n-dimensional vector, where each dimension corresponds to a term, or more precisely, to the TF-IDF weight of that particular term. To clarify this, let $M = \{m_1, m_2, m_3, ..., m_N\}$ be a set of visualizations and $T = \{t_1, t_2, t_3, ..., t_n\}$ a set of terms in M. Each visualization m_i is represented as a vector in a n-dimensional vector space, i.e., $m_i = w_{1,i}, w_{2,i}, w_{3,i}, ..., w_{n,i}$, where $w_{k,i}$ denotes the weight for the term t_k applied a visualization m_i, i.e.:

$$w_{k,i} = tf_{t_k,m_i} \times idf_t = tf_{t_k,m_i} \times \left[log_e\left(\frac{N}{df_t + 1}\right) + 1\right] \quad (1)$$

where the former factor of the product is an occurrence frequency of the term t_k applied a visualization m_i, and the later indicates the distribution of the term among the both profiles (i.e., so that particular and commonly occurring terms can be discriminated from each other). We apply the same weighting scheme to define the user profile. Having defined the profiles, it is now possible to estimate their similarity. To do so, we use the weighting information in the vectors and apply the *cosine similarity* measure [12], defined as follows:

$$sim(m_i, m_j) = \frac{\sum_k \mathbf{w}_{k,i}\mathbf{w}_{k,j}}{\sqrt{\sum_k (\mathbf{w}_{k,i})^2}\sqrt{\sum_k (\mathbf{w}_{k,j})^2}} \quad (2)$$

where m_j denotes the tag (or title, question) collection of the current user. The result of this measure is a cosine value of the angle between two vectors, in our case between the mapping combination and e.g., the tag collection. The retrieved values are then used as scores to rank the relevant visualizations following the Equation:

$$pred_{cb}(m_i, m_j) = \Sigma_{m_i, m_j \epsilon M} sim(m_i, m_j) \quad (3)$$

Note that the approach of our recommender system is described in detail in our previous paper and is beyond the scope of this paper. For more details please refer to [15].

4 EXPERIMENT SETUP

The goal of the study is to investigate the characteristics of tags, titles, and Q&As and their impact on recommending personalized visualizations. To collect these different kinds of annotations we designed a crowd-sourced study where we asked the user to annotate and rate the visualizations according to the different data sources. In Section 4.1, we provide details about how we collected the annotations (tags, titles, Q&As). In our previous work [16], we have already investigated the characteristics of the collected ratings and their impact on the recommendation quality— on our CF-RS respectively. Thus, here we put focus on tags, titles and Q&As. To that end, we proceed with the experiment as follows:

- First, we analyse how good these three types of input data models encode both user and visualizations (see Section 5). The observations from this part of the experiment shall reveal us some important facts about why some of the inputs are better than the other. Based on those observations, we build a list of candidate inputs for each data set (i.e., which descriptors accurately describe user, and which ones the visualizations). Those are in the end our assumptions that we want to confirm using the offline study.
- Next, we execute our content-based recommender on candidate input data models to see if their encoding power can be confirmed. The results of this study are presented in Section 6.

4.1 Datasets

Visualizations were generated for three open source datasets (see below) using a rule-based visualization recommendation system [15].

Note that the rule-based recommendation system uses heuristic rules that produce visually correct charts, but they are not always useful. Some examples of such charts are given in Figure 2, in the right column. They generally received low ratings since either they were visually useless or do not reveal much about the underlying data (cf. chart for the EU dataset) or they do reveal something, but not enough. For instance, the geo chart on bottom shows countries, but it actually hides all data about book publishers, which is essential to understand what is being visualized. Finally, there were also charts which show enough information, but have in fact received low ratings. These are typical cases where user expressed their subjective opinions (cf. stacked bar chart on top).

The following datasets have been used for the experiment:
MovieLens dataset (Movies): Top-ranked movies for the years 1960, 1970, 1980, and 1990. The rule based recommender produced for this dataset four types of visualizations (bar chart, line chart, timeline and geo chart) using the method described in [16] with the following frequencies: 32 bar charts, 9 line charts, 13 timelines and 1 geo-chart combinations. Hence, a total of 55 visualizations were generated.
EU Open Linked Data Portal dataset (EU): The percentage of the population looking for educational information online in the years 2009-2011 for 28 EU countries. The rule based recommender suggested 30 possible visualizations, concretely 15 bar charts, 6 line charts, 8 timelines and 1 geo chart.
Book-Crossing dataset (Books): 41 randomly chosen books published between 1960 and 2003. The rule based recommender suggested 3 visualization types: bar chart with 3 combinations, geo chart with 1 combination and timeline with 3 combinations, the total of 7 visualizations respectively.

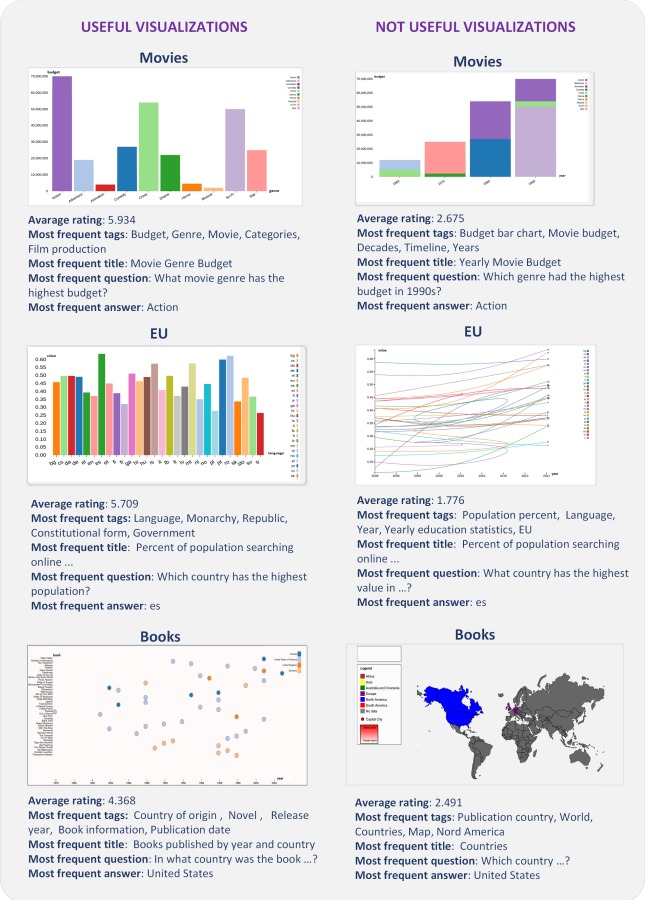

Figure 2: Examples of (in average) highest (left) and lowest (right) rated visualizations for all three datasets.

4.2 Procedure

A crowd-sourced experiment was carefully designed to obtain user preferences in different formats for each chart. While using a crowd-sourced platform, it is important to design the study so that participants do not blindly click through the options. Our datasources require ratings tags, titles, and Q&As for each chart. Following the suggestion of Kittur et al. [10], a cognitively demanding preparatory tasks should bring participants to accurately study the chart and prevent a random or rash answer. Therefore, the task (Human Intelligent Task, HIT) was designed as follows: a participant was given a one line description of a dataset originating the visualization, looking at the visualization she had to: 1) write tags (at most five), 2) write a title, 3) rate it, and finally 4) write a question the chart can answer. Figure 3 shows an example of a HIT. Rating a visualization with a single score would be rather unrealistic. Instead, a multidimensional rating scale lets the user consider various aspects of a visualization. We adapted a multidimensional scale from a list of usability factors presented in [24] and [29]. It included the following factors: (1) cluttered, (2) organized, (3) confusing, (4) easy to understand, (5) boring, (6) exciting, (7) useful, (8) effective and (9) satisfying. Note that dimensions 1-6 are duplicated with opposing sentiment (e.g., cluttered vs. organized). Opposing dimensions were used to ensure meaningful ratings for scales with complex meaning. Dimensions were rated on a 7-point Likert scale (1=not applicable 7=very applicable). After a pilot study, we decided to collect 3 charts per HIT, which makes for sensible task duration (5 min). Charts were distributed in 32 HITs, each with 3 randomly chosen charts. The procedure was as follows: After accepting a HIT, the participant (worker or turker) received a tour to complete a task, which showed a visualization and corresponding tags, title, ratings and Q&As in the exact same format as the subsequent study. The worker started the first task in the HIT by pressing a button. Workers were allowed to write not applicable (NA) for tags, title and Q&As, but were alerted if they failed to write any of these. The rating dimensions were not assigned a score until the worker did it. Workers could only proceed if they wrote the tags, a title, rated all dimensions and provided a question with the corresponding answer. A HIT with three charts was compensated with $1.00. A worker evaluated a minimum of three visualizations and was allowed to perform more than one HIT. Only expert workers who consistently achieved a high degree of accuracy by completing HITs were allowed to take part in the study. To make sure that the quality of collected data is satisfying, all entries have been manually verified. To that end, 10% of workers was rejected from the experiment, since they provided either incomplete or invalid inputs.

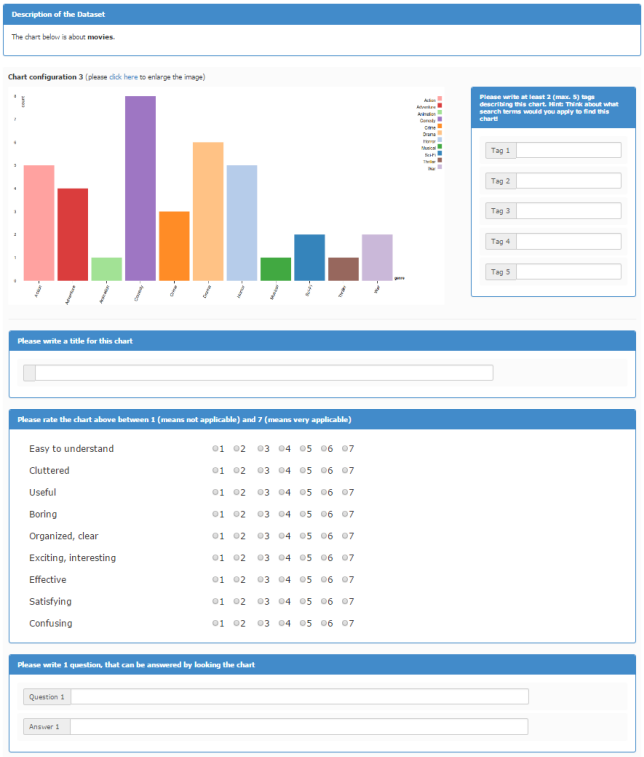

Figure 3: Example of a HIT for our crowd-sourcing experiment. Participants were motivated to carefully observe the visualization with the study task, in terms of writing tags and a title. Thereafter, they had to rate it in a multidimensional scale and pose a question that is answered with the visualization.

4.3 Participants

Each HIT was completed by ten workers. Note, this was the minimum required number on workers per visualization to train our recommender (see Section 6). In total, 47 workers (some of them were assigned to more than one HIT) completed our study. Workers completed on average 4.8 HITs. For 92 visualizations, we collected 8280 ratings across 9 dimensions, 4483 tags, 3881 titles and 4387 Q&As. The experiment started on November 26, 2014 and ended on December 3, 2014. The allotted working time per HIT was 900 sec and the average working time was 570 sec.

5 ENCODING POWER OF USER-PROVIDED INPUT

In this study we aim to explore the characteristics of different user-provided input (annotations) in terms of encoding users and visualizations. Information-theoretic measures are used to characterize the tags, titles, and Q&As.

5.1 Methodology

For the analysis of tags, titles and Q&As we employ information-theoretic measures: entropy, conditional entropy and mutual information. Using information-theoretic measures, we are able to (i) quantify the diversity in annotations (terms in further text), their encoding power respectively, and (ii) the amount of shared information between terms describing users and items (visualizations). With this information, we expect to answer why one input might be more suitable for recommending visualizations than the others.

In information theory, entropy measures the amount of uncertainty in a single random variable [6]. Given a random variable (X), which consist of occurrences $\{x_1 \cdots x_N\}$, each of which occurs with the probability $p(x)$, the entropy $H(X)$ is defined as:

$$H(X) = -\sum_{x \in X} p(x) log(p(x))$$

Conditional entropy [6], on the other side, measures the uncertainty in a random variable given the value of another random variable. Given two discrete random variables X = $\{x_1 \cdots x_N\}$ and Y = $\{y_1 \cdots y_N\}$ so that the event (x, y) occurs with the joint probability $p(x, y)$, the joint entropy is defined as:

$$H(Y,X) = -\sum_{\{y,x\} \in \{Y,X\}} p\,(y,x)\; log\,(p\,(y,x))$$

Using this value, conditional entropy is defined as $H(Y|X)$ [3]:

$$H(Y|X) = H(Y,X) - H(X)$$

Concretely, conditional entropy quantifies the amount of information needed to describe the variable X (e.g., user or visualization) when the value of the variable Y (e.g., tags, titles, Q&As) is known. If $H(Y|X)$ is minimized, each tag (or title, Q&A) uniquely refers to an individual user (or visualization) [3]. In contrast, when $H(Y|X)$ is maximized, each tag (or title, Q&A) is as likely as all others.

Finally, mutual information [6] is a measure of independence between two random variables. In other words, it quantifies the amount of data (information) shared (mutual) between variables. Given two discrete random variables X = $\{x_1 \cdots x_N\}$ and Y = $\{y_1 \cdots y_N\}$ so that the event (x, y) occurs with the joint probability $p(x, y)$, the mutual information $I(X;Y)$ is defined as:

$$I(X;Y) = H(Y) - H(Y|X)$$

High mutual information indicates a large dependency between two variables. In contrast, if the mutual information is minimized the variables are independent.

Table 1: Basic statistical properties of the datasets collected via Amazon Mechanical Turk. Column "User/Vis" shows the average number of user assigned to a visualization; "Vis/User" is the average number of visualizations assigned to a user. Note that the values in brackets indicate number of unique terms.

Dataset	#Vis	#User	Users/Vis	Vis/User	#Tags	#Titles	#Q&As
Movies	55	36	10	15.27	2731 (292)	2217 (295)	2638 (822)
EU	30	19	10	15.79	1403 (166)	1394 (234)	1354 (514)
Books	7	15	10	4.67	349 (87)	270 (92)	395 (188)

5.2 Results

Table 1 summarizes basic statistics for tags, titles and Q&As and shows the distribution of the entire terms [1] over user and visualizations. As introduced earlier, each of the visualizations in a

[1] A term is considered here as a single word e.g., in a tag input data model, a term corresponds to a single tag.

Table 2: Example distributions of top-5 terms for the EU dataset. Note, the terms are stemmed using Porter stemmer [9].

	Count	
Tags	# of visualizations	# of user
chart	20	4
countri	17	13
govern	14	8
onlin	14	8
valu	22	13
Titles		
constit	15	8
countri	22	5
european	14	8
popul	21	8
valu	17	8
Q&As		
inform	14	10
larg	16	9
onlin	13	9
republ	13	10
type	13	10

particular dataset was individually evaluated regarding to tags, titles, and Q&As, i.e., 55 visualizations in Movies dataset, 30 and 7 in EU and Books respectively (cf. the second column in Table 1). For this configuration, user involved in the study have provided overall 4483 tags, 3881 titles, and 4387 Q&As (2% yes/no Q&As). The average worth length (char) was 5.2 for tags and questions, 5.3 for titles, and 4.7 for answers. An excerpt of the most popular terms for EU is shown in Table 2. Some important differences between collected data could already be identified when considering this distribution in conjunction with unique terms (Note that unique terms are enclosed with brackets, see Table 1). According to descriptive data from the table, 10.69% of the tags, 13.31% of titles and 31.16% of Q&As were unique, i.e., not globally repeated. The fact that a question typically associates with only one specific visualization may explain this phenomenon. Taking this cue, we can assume that the varied number on different type of terms directly affects the recommendation quality. In brief, the more unique terms are applied to a visualization the easier it should become to discriminate this visualization in the finding process from others. Subsequently, the more individual terms a user provides, the higher the ability should be to accurately direct this user to the preferred visualizations [25]. However, the more accurate way to measure how good a term is in discriminating a resource from others is measuring the value of the information it provides about a resource and about the user. For this purposes we investigate in the following, first, the power of users' terms at encoding users' visual preferences, and, next, at encoding the content of visualizations.

5.2.1 Power of user-provided input at encoding users. To investigate the quality of extracted terms at encoding users' visual preferences we applied the information-theoretic measures among all three datasets. Considering all three datasets in our analysis helps us to achieve more objective results, compared to analyzing each dataset individually.

Table 3: Information-theoretic measures for tags, titles and questions used for user profiles. Note that the measures have been calculated among all three datasets.

		User Model		
Datasets	**Term**	**Entropy**	**Conditional Entropy**	**Mutual Inf.**
Movies, EU, Books	Tags	5.9376	3.0381	2.8995
	Titles	6.1421	2.9815	3.1606
	Questions	6.8898	3.1436	3.7462

In this experiment, X is users and Y is either tags, titles or Q&As. The analysis intends to determine which of $H(Tags)$, $H(Titles)$, $H(Q\&As)$ indicates more diversity, which of $H(User|Tags)$, $H(User|Titles)$, $H(User|Q\&As)$ has more power in describing users, and which of $I(User;Tags)$, $I(User;Titles)$, $I(User;Q\&As)$ has higher value and can specify users better. Table 3 summarizes the results of this study. Note, to follow a common design principle of interactive (question-answering) systems, we suggest to split the Q&As input so that questions are used for the user- and answers for the item model.

When considering the results in Table 3, at the first look we can observe that the entropy ($H(Questions)$) is higher than ($H(Titles)$) and ($H(Tags)$). This suggests, users provided more diverse and specific questions than titles and tags. Given this fact, we hypothesize that questions have a strong encoding power. Yet, entropy measures the amount of uncertainty. Conditional entropy, however, quantifies the amount of uncertainty in a random variable (i.e., user) given the value of another random variable (i.e., tags, titles or questions).

We therefore consider next the entropy of users conditional on tags (or titles, questions), i.e., $H(User|Tags)$, $H(User|Titles)$ and $H(User|Questions)$ (see Table 3 second column). Looking at the results, $H(User|Questions) > H(User|Tags) > H(User|Titles)$. What that means is, that tags and titles have a strong power in describing user than questions.

Yet, conditional entropy is a relative measure and tells little about the independence between tags (or titles, questions) and user [3]. The independence, however, matters in recommender systems when it comes to defining a link between user and resources. Thus, to complete the analysis on tags (or titles, questions) in specifying user, we next, analyze the amount of information shared between tags (or titles, questions) and a user, mutual information ($I(User;Tags)$, $I(User;Titles)$, $I(User;Questions)$) respectively. The results show that $I(User;Questions)$ is the highest compared to $I(User;Tags)$ and $I(User;Titles)$ (see Table 3 last column). Yet, these results finally suggest, questions are more effective in specifying user than tags and titles.

5.2.2 Power of user-provided input at encoding visualizations. Similar to our previous study, to investigate the general quality of the user-provided input at encoding visualizations we applied the information-theoretic measures among all three datasets. In this case, X are visualizations and Y are either tags, titles or answers. The analysis intends to determine which of $H(Tags)$, $H(Titles)$,

Table 4: Information-theoretic measures for tags, titles and answers used for item profiles. Note that the measures have been calculated among all three datasets.

Datasets	Term	Item Model		
		Entropy	Conditional Entropy	Mutual Inf.
Movies, EU, Books	Tags	5.9376	4.1429	1.7947
	Titles	6.1421	4.1384	2.0037
	Answers	6.6371	2.7405	3.8966

$H(Answers)$ indicates more diversity across visualizations, which of $H(Vis|Tags)$, $H(Vis|Titles)$, $H(Vis|Answers)$ has more power in describing visualizations, and which of $I(Vis;Tags)$, $I(Vis;Titles)$, $I(Vis;Answers)$ has higher value and can specify visualizations better. Table 4 summarizes the results of this study. The entropy of answers is higher than of tags and titles. At a first glance, this indicates, the visualizations have been annotated with more specific and unique answers than tags and titles. However, as we noted in the previous study, entropy just measures the amount of uncertainty in a random variable (i.e., visualization) given the value of another random variable (i.e., tags, titles or answers). When considering $H(Vis|Tags)$, $H(Vis|Titles)$, $H(Vis|Answers)$, we observe that answers are more unique and special than tags and titles ($H(Vis|Answers) < H(Vis|Tags)$, $H(Vis|Answers) < H(Vis|Titles)$ (see Table 4 second column). Thus, it might be more difficult for the system to retrieve a visualization that has been annotated with a certain tag or title than with a certain answer. To validate this we finally measure the degree of independence between tags (or titles, answers) and a visualization– the amount of information shared (mutual) $I(Vis;Tags)$, $I(Vis;Titles)$, $I(Vis;Answers)$. Remember, full independence is reached when e.g., $I(Vis;Tags)$ is zero.

Table 4 (last column) shows the mutual information $I(Vis;Answers)$ is higher than of $I(Vis;Tags)$ and $I(Vis;Titles)$. These results, finally, suggest a high quality of answers at encoding visualizations. Taking this cue, we can assume the answers are powerful to direct the user to the corresponding visualizations than tags and titles.

5.2.3 Summary. Using information-theoretic measures we aimed to characterize tags, titles and Q&As in describing user and items (visualizations). To that end we performed two studies where we analyzed the power of (i) tags, titles and questions at encoding user, and (ii) tags, titles and answers at encoding visualizations. The findings of the studies should help in predicting performance of the potential candidates for the user- and item models being used for our visual recommender.

Results suggest a strong link (dependency) between user and her questions and items and their (assigned) answers. This assumption is made regarding to the shared information between (i) user & questions, and (ii) item & answers,
$I(User;Questions)$, $I(Vis;Answers)$ respectively. Namely, the results of $I(User;Questions)$, $I(Vis;Answers)$ show that a set of specific terms from questions refers to an individual user and each answer to a specific item. Yet, this is an essential finding for designer of content-based recommender systems. It suggests using questions for user modeling and answers for the item modeling.

To verify this assumption, we build, in the following, user and item models using user's questions and answers and explore the quality of the generated recommendations in an offline study employing our CB based recommender system. We applied this recommender technique since it is traditionally used for user-provided input, such as tags, comments, etc. We measured the quality of the recommendations by their closeness to what user prefers and needs.

Note, for the sake of completeness, we also included additional setting where tags are taken for user- and item models. Considering the results in Section 5.2.1, the quality of the generated recommendations should be lower when using this combinations, since tags have a lower mutual information than Q&As. Moreover, to verify the low performance of titles, settings with titles are reported too.

In the following we describe the method and metrics used to validate our approach in detail and present the results of the offline study.

6 RECOMMENDATION QUALITY

6.1 Methodology

Following the method described in [15], we split the preference model including either users' tags, titles or questions into the two distinct sets, one for training the recommender (training-set), and another one for testing (test-set). The test-set acts here as a reference value that, in an ideal case, has to be fully predicted for the given training-set. From each of the datasets in the preference model, we randomly select 80% of user's data and enter them into the training-set performing 5-fold cross validation. The recommendations produced out of the training-set are further used to evaluate the performance of our recommender. The performance of the recommender depends generally on how good it predicts the test-set. We compared the generated recommendations (prediction-set) and the test-set by applying a variety of well-known evaluation metrics in information retrieval [7]: Recall (R), Precision (P), F-Measure (F), Mean Average Precision (MAP) and the Normalized Discounted Cumulative Gain (nDCG). The first three metrics basically express the quantity of relevant recommended results, whereas MAP and nDCG quantify the concrete ordering of the results (i.e., penalizing results which are not on the top but are relevant for the user). We refer to the research papers [19, 21, 23] for more detailed definitions of the evaluation metrics. Note, the tests are performed for each user- and item model combination independently.

6.2 Results

To measure the improvements in terms of recommender quality (=accuracy, relevance), we compared the individual CBs ($CB_{Tags,Tags}$, $CB_{Titles,Titles}$, $CB_{Q,A}$) with the baseline filtering algorithm Random (RD). The RD method simulates the recommender behavior providing a random rating for each visualization. Note, for the Q&As based CB approach ($CB_{Q,A}$) we used user's questions in user- and user's answers in item model.

For the comparison, we analyzed the top 3 recommendations (k=3), since our datasets are relatively smaller than some commonly used datasets, such as CiteULike and BibSonomy. Table 5 shows the quality metrics values F@3, MAP@3, nDCG@3 estimated for the three datasets.

Table 5: The performance of our individual content based filtering approaches (CB), compared with baseline algorithm RD: quality metric values considering the first three recommendations in the list (k=3). ***Significant at p<0.001.

		Metric		
Dataset	**Algorithms**	**F@3**	**MAP@3**	**nDCG@3**
Movies	**RD**	0.0055	0.0020	0.0048
	$\mathbf{CB}_{Tags,Tags}$	0.0740***	0.0545***	0.0830***
	$\mathbf{CB}_{Titles,Titles}$	0.0650***	0.0500***	0.0743***
	$\mathbf{CB}_{Q,A}$	0.0547***	0.0450***	0.0643***
EU	**RD**	0.0150	0.0044	0.0103
	$\mathbf{CB}_{Tags,Tags}$	0.1862***	0.1120***	0.1801***
	$\mathbf{CB}_{Titles,Titles}$	0.1726***	0.1030***	0.1663***
	$\mathbf{CB}_{Q,A}$	0.1505***	0.1014***	0.1642***
Books	**RD**	0.0333	0.0333	0.0420
	$\mathbf{CB}_{Tags,Tags}$	0.2360***	0.2077***	0.2700***
	$\mathbf{CB}_{Titles,Titles}$	0.2310***	0.2133***	0.2720***
	$\mathbf{CB}_{Q,A}$	0.2267***	0.2233***	0.2720***

Yet, when considering the recommendation accuracy (F@3), at a first glance, we can observe that tags based CB ($CB_{Tags,Tags}$) outperforms for all three datasets the baseline algorithms RD (cf. F@3($CB_{Tags,Tags}$) = 0.0740, F@3(RD) = 0.0055 for Movies). So, we hypothesize that the experimentation with individual user- and item models has had some effect among all three datasets. To discover what the effect was and how significant it is, we performed statistical tests which we report in the following.

The results for F@3, MAP@3, nDCG@3 have been analyzed independently for each dataset applying Friedman's ANOVA. Note, we used this test since our data were not normally distributed and (per dataset) the same participants have been used for each individual CB approach. The results for all three datasets show a significant effect of the used type of item- and user models on the recommendation accuracy (F@3), with $\chi^2(4) = 25.10$ for Movies, $\chi^2(4) = 19.80$ for EU, and $\chi^2(4) = 20.14$ for Books, $p<0.001$. To explore where the differences lie we applied *Post hoc* tests with Bonferroni correction. The results for all three datasets reveal a significant difference between the values of the individual CB approaches ($CB_{Tags,Tags}$, $CB_{Titles,Titles}$, $CB_{Q,A}$) and baseline algorithm. Note, the critical difference ($\alpha = 0.05$ corrected for the number of tests) was 28.10 for Movies, 20.10 for EU and 18.65 for Books. However, there were no significant differences between the values of individual recommenders ($CB_{Tags,Tags}$, $CB_{Titles,Titles}$, $CB_{Q,A}$), $p>0.05$.

Looking at the results for MAP@3 and nDCG@3 measures which examine the ranking of the recommended visualizations we observe similar results. Concretely, the results show a significant effect of the used type of item- and user models on the ranking of the recommendations, with $\chi^2(4) = 23.56$ for Movies, $\chi^2(4) = 20.10$ for EU, and $\chi^2(4) = 18.65$ for Books, $p<0.001$. Similar to the previous analysis, to explore where the differences lie we applied *Post hoc* tests with Bonferroni correction. The results of *Post hoc* tests shown for all three datasets, when tags, title and Q&As based models are used, the visual recommender can sort the recommendations according to their relevance better than baseline algorithm. The critical difference ($\alpha = 0.05$ corrected for the number of tests) was 28.10 for Movies, 20.10 for EU and 18.65 for Books. The results for nDCG confirmed the results we obtained for MAP@3 measures showing a significant improvement by ranking of recommendations when using either (i) tags, (ii) titles or (ii) Q&As based models compared to the random baseline algorithm.

7 DISCUSSION

The main outcome of our study is that all three inputs (tags, title, and Q&As) show a comparable quality in recommending visualizations. This result is important because it gives the designer freedom in choosing the method for preference elicitation. Besides, it makes the suggested approach applicable in domains in which only particular types of inputs can be supported (e.g., question-answering systems).

We could confirm this result for all datasets, as illustrated in Table 5. Moreover, all three inputs are, as expected, significantly better at encoding visualizations than the baseline algorithm RD. Also, when considering the results in more detail, i.e., the quality *F*@3, and the sorting accuracy (*MAP*@3 and *nDCG*@3), it does not matter which of the inputs to use. (Note that there are negligible differences in means, which are statistically not significant). This would, in the end, mean that characteristics of the individual inputs are very close to each other. In fact, providing a title would be nothing else but providing a set of tags (in terms of how many and which words have been provided). We analyzed these characteristics in the first part of our study.

Using information-theoretic measures we found that some inputs better encode user/visualizations than the other. In particular, questions and answers have been identified to show distinctive characteristics compared to tags and titles. It turned out that they more precisely address a particular user/visualization, since, as results reveal, they have terms which are less common (shared) than in the case of tags and titles. This, in fact, comes from the nature on how questions/answers are built. For instance, it is more likely that similar or same words are provided when describing visualizations via tags rather than using complex sentences. Generally, users are familiar when describing resources in form of tags, as tagging approach is quite intuitive and straightforward. Using question/answers, instead, is more subjective. One aspect here is building a sequence of words (a sentence), and another is using proper adjectives in that sequence. These terms also contribute to the user/item model. Nevertheless, as shown later in the offline experiment, these differences were not significant enough to be manifested by the content-based recommender (at least with the cosine similarity metric we chose).

8 CONCLUSIONS

In this paper we investigated the power of different kinds of user-provided input to effectively encode user's visual preferences and the content of visualizations. To do so we employed information-theoretic measures including entropy, conditional entropy and mutual information. Using these measures, we were able to quantify the diversity in individual inputs, their encoding power respectively, and also the amount of shared information between them and users/visualizations. The outcome of the study should suggest a list of potential candidates to build user models defining users' interest/needs and item models describing the content of

the visualization— both crucial for content-based recommender systems. Finally, we executed our content-based recommender on candidate models to see if their encoding power could be confirmed. In other words, we performed an offline study to assess the practicability of the individual models in recommending personalized visualizations. The data we used in this paper was collected in the scope of the empirical study, where we involved 47 participants to annotate different types of visualizations using tags, titles, questions and answers.

Regarding to our first study, we found that the best user- item model combination is guaranteed when using questions for the user- and answers for the item models (considering their mutual information values). The offline study has confirmed the good quality of this combination as it produced better recommendations than the baseline algorithm. However, the quality of this combination was not significantly better or different than that of the tags and titles. Although differences at encoding power between the individual inputs could be manifested, those differences were negligible and not crucial for the content-based recommender system we employed. Nevertheless, the fact that the recommendation quality and accuracy were still high using the alternative inputs, titles and Q&As respectively, demonstrated the capability of these inputs being used for visual recommender systems.

In summary, this paper shows the good quality of alternative input types (titles, Q&As) to derive high quality visualization recommendations. It further emphasizes the relevance of annotations for the users as they directly link them to the items which might be closer to what they need and prefer.

Our research so far did not concentrate how a hybrid recommender would perform when using user's ratings with titles or Q&As. This is planned for the near future. In the current work, we used each information sources separately. In the future, we will investigate how our CB performs when using a combination of multiple information sources as data model. Furthermore, we plan to investigate interfaces to elicit such information with minimal effort making it part of the analysis process whenever possible.

9 ACKNOWLEDGMENTS

This work is funded by the European Horizon 2020 research project AFEL (grant nr. 687916) and CONICET. The Know-Center GmbH is funded within the Austrian COMET Program - managed by the Austrian Research Promotion Agency (FFG).

REFERENCES

[1] Alejandro Bellogín, Iván Cantador, and Pablo Castells. 2010. A Study of Heterogeneity in Recommendations for a Social Music Service. In *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*. ACM, 1–8.

[2] Toine Bogers and Antal Van den Bosch. 2009. Collaborative and content-based filtering for item recommendation on social bookmarking websites. In *Proceedings of the ACM Recommender Systems workshop on Recommender Systems and the Social Web*, Vol. 9. 9–16.

[3] Ed H. Chi and Todd Mytkowicz. 2008. Understanding the Efficiency of Social Tagging Systems Using Information Theory. In *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*. ACM, 81–88.

[4] Micheline Elias and Anastasia Bezerianos. 2012. Annotating BI Visualization Dashboards: Needs & Challenges. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1641–1650.

[5] Umer Farooq, Yang Song, John M. Carroll, and C. Lee Giles. 2007. Social Bookmarking for Scholarly Digital Libraries. *Internet Computing, IEEE* 11, 6 (05 Nov. 2007), 29–35.

[6] Robert M. Gray. 1990. *Entropy and Information Theory*. Springer-Verlag New York, Inc.

[7] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems* 22, 1 (Jan. 2004), 5–53.

[8] Jiwoon Jeon and R. Manmatha. 2004. *Using Maximum Entropy for Automatic Image Annotation*. Springer Berlin Heidelberg, 24–32.

[9] Wahiba Ben Abdessalem Karaa and Nidhal Gribâa. 2013. Information Retrieval with Porter Stemmer: A New Version for English. In *Advances in Computational Science, Engineering and Information Technology*. Vol. 225. Springer International Publishing, 243–254.

[10] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 453–456.

[11] Yi-Ling Lin, Christoph Trattner, Peter Brusilovsky, and Daqing He. 2015. The impact of image descriptions on user tagging behavior: A study of the nature and functionality of crowdsourced tags. *Journal of the Association for Information Science and Technology* 66, 9 (2015), 1785–1798.

[12] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2011. Content-based Recommender Systems: State of the Art and Trends. In *Recommender Systems Handbook*. Springer US, 73–105.

[13] Jared Lorince, Sam Zorowitz, Jaimie Murdock, and Peter M. Todd. 2015. The Wisdom of the Few? "Supertaggers" in Collaborative Tagging Systems. *Journal of Web Science* 1, 1 (2015), 16–32.

[14] Qiaozhu Mei and Kenneth Church. 2008. Entropy of Search Logs: How Hard is Search? With Personalization? With Backoff?. In *Proceedings of the International Conference on Web Search and Data Mining*. ACM, 45–54.

[15] Belgin Mutlu, Eduardo Veas, and Christoph Trattner. 2016. VizRec: Recommending Personalized Visualizations. *ACM Transactions on Interactive Intelligent Systems* 6, 4, Article 31 (Nov. 2016), 31:1–31:39 pages.

[16] Belgin Mutlu, Eduardo Veas, Christoph Trattner, and Vedran Sabol. 2015. Towards a Recommender Engine for Personalized Visualizations. In *User Modeling, Adaptation and Personalization*. Vol. 9146. Springer International Publishing, 169–182.

[17] F. Matsatsinis Nikolaos, Lakiotaki Kleanthi, and Tsoukiás Alexis. 2011. Multicriteria User Modeling in Recommender Systems. *IEEE Intelligent Systems* 26 (2011), 64–76.

[18] Thomas Orgel, Martin Höffernig, Werner Bailer, and Silvia Russegger. 2015. A metadata model and mapping approach for facilitating access to heterogeneous cultural heritage assets. *International Journal on Digital Libraries* 15, 2-4 (2015), 189–207.

[19] Denis Parra and Shaghayegh Sahebi. 2013. Recommender Systems: Sources of Knowledge and Evaluation Metrics. In *Advanced Techniques in Web Intelligence-2*. Vol. 452. Springer Berlin Heidelberg, 149–175.

[20] Erhard Rahm and Philip A. Bernstein. 2001. A Survey of Approaches to Automatic Schema Matching. *The VLDB Journal* 10, 4 (Dec. 2001), 334–350.

[21] Majdi Rawashdeh, Heung-Nam Kim, JihadMohamad Alja'am, and Abdulmotaleb El Saddik. 2013. Folksonomy link prediction based on a tripartite graph for tag recommendation. *Journal of Intelligent Information Systems* 40, 2 (2013), 307–325.

[22] Francesco Ricci and Quang Nhat Nguyen. 2007. Acquiring and Revising Preferences in a Critique-Based Mobile Recommender System. *IEEE Intelligent Systems* 22, 3 (May 2007), 22–29.

[23] C.J. Van Rijsbergen. 1974. Foundation of Evaluation. *Journal of Documentation* 30, 4 (1974), 365–373.

[24] Ahmed Seffah, Mohammad Donyaee, Rex B. Kline, and Harkirat K. Padda. 2006. Usability Measurement and Metrics: A Consolidated Model. *Software Quality Control* 14, 2 (June 2006), 159–178.

[25] Markus Strohmaier, Christian Koerner, and Roman Kern. 2010. Why Do Users Tag? Detecting Users' Motivation for Tagging in Social Tagging Systems. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*.

[26] Fernanda B. Viegas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. 2007. ManyEyes: A Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (Nov. 2007), 1121–1128.

[27] William Wright, David Schroh, Pascale Proulx, Alex Skaburskis, and Brian Cort. 2006. The Sandbox for Analysis: Concepts and Methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 801–810.

[28] Y. Y. Yao. 2003. *Information-Theoretic Measures for Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, 115–136.

[29] Xianjun Sam Zheng, James j W. Lin, Salome Zapf, and Claus Knapheide. 2007. Visualizing User Experience Through "Perceptual Maps": Concurrent Assessment of Perceived Usability and Subjective Appearance in Car Infotainment Systems. In *Proceedings of the 1st International Conference on Digital Human Modeling*. Springer-Verlag, 536–545.