Enhanced User Interaction to Qualify Web Resources by the Example of Tag Rating in Folksonomies

Monika Steinberg, Orhan Sarioglu, Jürgen Brehm
Institute of Systems Engineering - System and Computer Architecture
Hannover, Germany
[steinberg, brehm]@sra.uni-hannover.de, orhansarioglu@freenet.de

Abstract - **The Web offers autonomous and frequently useful resources in growing manner. User Generated Content (UGC) like Wikis, Weblogs or Webfeeds often do not have one responsible authorship or declared experts who checked the created content for e.g., accuracy, availability, objectivity or reputation. The user is not able easily, to control the quality of the content he receives. If we want to utilize the distributed information flood as a linked knowledge base for higher-layered applications – e.g., for knowledge transfer and learning – information quality (iq) is a very important and complex aspect to analyze, personalize and annotate resources [1]. In general, low information quality is one of the main discriminators of data sources on the Web [2]. Assessing information quality with measurable terms can offer a personalized and smart view on a broad, global knowledge base. We developed the qKAI application framework [3] to utilize available, distributed data sets in a practically manner. In the following, we present our adaption of information quality aspects to qualify Web resources based on a three-level assessment model. We deploy knowledge-related iq-criteria as tool to implement iq-mechanisms stepwise into the qKAI framework. Here, we exemplify selected criteria of information quality in qKAI like relevance or accuracy. We derived assessment methods for certain iq-criteria enabling rich, game-based user interaction and semantic resource annotation. Open Content is embedded into knowledge games to increase the users' access and learning motivation. As side effect the resources' quality is enhanced stepwise by ongoing user interaction. By the example of image tag rating in folksonomies we demonstrate a practicable use case for qualifying web resources by keyword-oriented group search and game-based tag ranking in detail.**

Keywords - Information Quality, Folksonomy, Open Content, Semantic Annotation, Knowledge Transfer.

I. INTRODUCTION

If we want to embed Web content into knowledge transfer and learning, the question about the data's' quality is indispensable. Information quality (iq) is an important concern if we want to build knowledge out of information towards education. Currently, Web users are claiming for more sophisticated content and less triviality [4]. To utilize autonomous web resources qualitative assessment of the broad information load becomes more and more important.

To let users interact with Open Content [5] out of distributed web resources, enhanced inquiry, selection, storage and buffering are important prerequisites. Nevertheless, statements of the resources iq enhance its fitness for use. The more we know about a resource, the better we can reuse it. We developed the qKAI application framework (qualifying Knowledge Acquisition and Inquiry) [3] - a service-oriented, generic and hybrid approach combining knowledge related offers for convenient reuse. As part of the qKAI application framework, we implemented the qKAI hybrid data layer to acquire, store and represent Open Content out of distributed resources. In qKAI Open Content is boosted as an inherent part of higher-layered applications in knowledge and information transfer via standard tasks of knowledge engineering and augmented user interaction. Especially regarding smart user interaction, we have to offer user interfaces with high scores in certain information quality criteria. If we get to know about a resource that it contains Chinese text by analyzing its metadata, we can deduce that its "*understandability*" is almost not ideal for European users. There are lots of small hints and tasks that are very helpful altogether to assess and enhance information quality aspects of Open Content. In the following, we introduce the meaning of information quality in qKAI exemplified with selected criteria of iq. We explain the relation between these criteria, qKAI data and interaction issues with Open Content.

A. Structure of this contribution

First, we introduce some further background. In Section 2 follows the state of the art. Section 3 gives an overview of information quality (iq) criteria. Section 4 offers some more details about assessing Web contents' quality. Section 5 shows how to combine traditional information quality metrics with enhanced user interaction, Section 6 exemplifies the use case of pictures' relevance in folksonomies, Section 7 gives a short analyzes of tagging systems. Section 8 explains our derived approach: keyword-oriented group search for tag ranking in folksonomies. Section 9 shows our game-based tag rating approach qRANK. Section 10 illustrates some evaluation results regarding our approaches versus the standard Flickr keyword search. Section 11 offers further application scenarios and use cases. At least this contribution ends up with conclusion and outlook in Section 12.

B. Utilizing Open Content for knowledge transfer and learning

The qKAI application framework serves as basis to develop rich user interaction with Open Content [6] upon it. Actually, we implement and evaluate knowledge visualization and game prototypes. qMAP acts as an interface to visualize and interact with Open Content like images, texts or videos geographically on a map. qMATCH offers the user term-image or term-term assignment questions out of Flickr content. qCHUNK lets the user guess Wikipedia articles while he gets presented chunks out of them. qMAP is a geocoded, map-based gaming board to visualize Open Content like Wikipedia articles or Flickr images. Some examples are shortly presented in Chapter 6. We see game-based interaction as a use-case with high design and interaction receivables that is well suited to evaluate enhanced interaction with Open Content exemplary.

C. Assessing the information quality of autonomous web resources

"*Information quality (iq) is one of the main discriminators of data and data sources on the Web. ... The autonomy of Web data sources renders it necessary and useful to consider their quality when accessing them and integrating their data.*" [2].

Information quality is often described as "*fitness for use*" [7] in the relevant literature. Metadata plays an important role for the determination of iq-criteria. Information quality is to a great extend subjective, because we have to mention multi-dimensional criteria while assessing context-, user- and task-dependent. Subjective dimensions of iq must be assessed by the help of user interaction [2]. User interaction can be basic, direct or indirect feedback.

… "*Many iq-criteria are of subjective nature and can therefore not be assessed automatically, i.e., independently and without help of the user.*"… [2]

Because iq is often subjective, task- and context-dependent, **user interaction** plays a very important role while assessing subjective iq-criteria. To let users rate and rank content according to certain iq-criteria, questionnaires are widely used.

D. Semantic annotation of resources

qKAI delivers an URI about every resource it utilizes in RDF [8] representation. Semantic interlinking between the provenance resource and the new, annotating qKAI URI connects the URIs following Linked Data paradigms. Semantic interlinking allows following all references (links) automatically. HTML for example does not offer this ability.

E. A global interaction rewarding model (GIAR)

An ontology-based interaction rewarding model (GIAR) is work in progress. qKAI rewards any kind of interaction with resources and other users to increase user participation and incentive. Therefore, we are designing a catalogue of interaction tasks and order them according to domain, type and further iq- criteria. For every interaction the user earns points according to a global point and level system like in game-based scenarios. We are rewarding external activity also, like e.g., listening to music at Last.fm, making friends at Facebook or tweets at Twitter. Every interaction is stored in a personal profile file that builds knowledge-related reputation step by step. Every resource has its own transaction and interaction protocol (see Figure 3 in Chapter 6). The protocol can be statistically evaluated to enable automated ranking, rating and deriving further iq-criteria. A social interaction rewarding community is under development to visualize the global interaction rewarding concept.

II. State of the Art and Related Work

Wang [9], Naumann [2] and Bizer [13] a.o. offer comprehensive research work about categorization, definition of information quality and related vocabulary in the domain of webbased information system. Wikipedia [10] has its own quality assessment deploying a review mode by authors. Freebase [11] allows the user to rearrange, connect, correct or annotate available resources. Rating, ranking and recommendation at Amazon [12] are good examples for enhanced user interaction to qualify content. Flickr offers properties related to a picture that enable to rate a photos quality. Tagging allows users to restructure and weight their knowledge in a self-controlled way. Revyu [14] allows the users to rank and rate everything. In qKAI we will integrate Revyu by querying whether a resource is annotated by Revyu yet. The reputation of a thing, person or resource in qKAI is increased if there is a Revyu entry about it. The existence of available interlinked context information in e.g., other web platforms is a first and simple step to determine information quality of resources according to scores.

III. Information Quality Criteria and Open Web Content

The "*fitness for use*" can depend on numerous factors like actuality, believability, completeness or relevance. Not all single criteria are assessable independent from each other [13]. Next to several further properties the most important criteria of information quality in web applications are actuality, reputation, believability and accuracy of content.

In contrast to processes inside of enclosed organizations that analyze iq as cyclic management task the assessment of iq in the Web relies on autonomous information providers in an open information space. Therefore, in webbased systems IQ is assessed by the help of user interaction to determine the "*fitness for use*" of an information source for the specific task on hand [13]. Social aspects of iq especially in the context of Web 2.0 are reputation and trust of the author.

Important for the believability of information is the reputation of the creator. Every user has his own opinion based upon own experience or the experience in his knowledge circle. All experiences that are made with resources in qKAI are logged in history protocols. Different opinions about the reliability or trustworthiness of single actors regarding certain themes emerge. Personalized knowledge views can be deduced this way.

There are trust metrics and policies for **reputation-based systems** available in literature and research [15] that can be implemented next to **interaction-based** and **metadata-relying** metrics.

A. Categorizing Information Quality

The categorization of information quality is in respective literature available according to various criteria and dimensions [16]. We did not find much about generic interaction components to assess ongoing iq in web-based knowledge systems by online assessment [17] components with game-based features. We see especially the combination of reputation-based and global metrics as promising first step towards an incentive and motivating way to assess iq sustainable.

TABLE I. IQ CITERIA AND THEIR CLASSIFICATION FOR AUTONOMOUS INFORMATION SYSTEMS BASED ON C. BIZER'S CATEGORIZATION [13]

Category	Criteria/Dimension	Objective/ subjective
Intrinsic criteria (Independent of the user's context)	Accuracy*	objective
	Consistency	objective
	Objectivity	objective
	Timeliness	objective
Contextual criteria (Context, task and user dependent)	Believability	subjective
	Completeness	subjective
	Understandability	subjective
	Relevancy	subjective
	Reputation	subjective
	Verifiability	subjective
	Amount of Data	subjective
Representational criteria	Interpretability	subjective
	Rep. Conciseness	subjective
	Rep. Consistency	objective
Accessibility criteria	Availability	objective
	Response Time	objective
	Security	objective

*Accuracy is interpreted in a bias way in qKAI: On the one side, we have to assess the data accuracy, on the other side we speak of semantically and syntactically correct information. The last one can only be assessed by enhanced user interaction of experts or collective intelligence approaches (Wisdom of crowds).

Accuracy is defined as the percentage of data without data errors, such as non unique keys or out of range values. Mohan et al. give a list of possible data errors [2].

B. Iq-criteria for the qKAI system domain

It is not practicable to measure all available iq-criteria at once. We have to select the most important criteria for our domain. In qKAI we have a strong focus on knowledge transfer with smart interaction. To offer knowledge-related content, we have to fulfill e.g., semantically correctness of factual data. We interpret semantically correctness as one aspect of accuracy. Accuracy is defined as the degree of correctness and precision with which information in an information system represents states of the real world [14]. **Figure 1** shows the actually most important iq-criteria in the qKAI system domain.

Technical or also called accessibility criteria like availability, response time or security depend almost on soft- and hardware concerns. We developed the qKAI hybrid data layer as part of the qKAI application framework to offer good results for these technically oriented criteria on an affordable Quadcore-platform. qKAI is suitable to search and explore distributed resources in an effective manner and represents our ongoing and enhanced research toward hybrid data management for distributed resources with rich interaction on top of it. To reach good results in the frontend, the backend – including the data layer - has to be suitable for this purpose. E.g., if a user waits too long, to get first search results, the motivation to ongoing interaction will rapidly increase. The iq-criteria "*response time*" and "*availability*" have to be enhanced by technically aspects like hard- or software requirements.

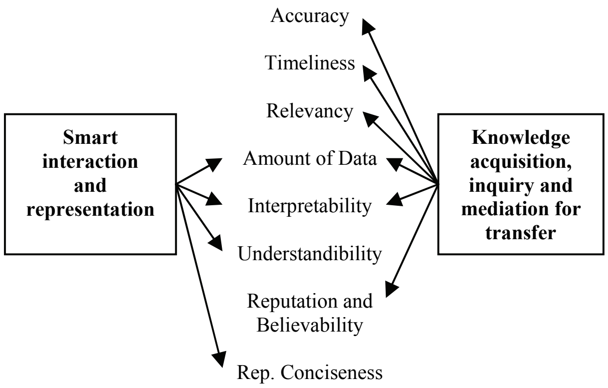

Figure 1. Most relevant iq-criteria for the qKAI system domain: knowledge transfer and smart interaction based on autonomous resources

C. Reputation as quality criteria and for users' motivation

Reputation can be seen as the sum of single experiences and expectation about trustworthiness and competence of a person, a group or an organization. Reputation has much to do with image and status of a person or a thing and is an important factor in online communities, where trust and reliability come into play. Most online communities that collect feedback to qualify content do not offer any incentive to rate and rank. The missing motivation of users to interact on the content is an essential problem, because there are no rational reasons to participate sustainably and the chance is taken to let other users do the ratings [18]. Creating and enhancing the own reputation is next to the simple fun [19] a good motivator to embed online users into to content-related participation without material incentive [18] [13]. Ebay and Amazon are successful examples for building reputation by users' feedback. In qKAI, the reputation of users is stored implicitly in their personal profile and increases with every kind of interaction on Open Content. A resources reputation is stored in their semantically linked qKAI annotation URI and is also increased by any interaction or analyzes, the resource is involved.

IV. ASSESSING THE QUALITY OF WEB CONTENT

…"*Information quality assessment is the process of assigning numerical values (iq-scores) to iq-criteria. An iq-score reflects one aspect of information quality of a set of data items.*" …[1]

To assess the iq of information sources, a scoring function calculates assessment scores form the collected ratings. The scoring function decides which ratings are taken into account and might assign different weights to ratings. Which criteria to take for a specific rating should be adjustable by the user and his task on hand? Our research showed the following classifications and assessment models to be most suitable for qKAI and webbased information systems with knowledge-related concerns in general. Naumann identified three main factors the quality of information is influenced by in his query-oriented approach:

- the perception of the user (the subject of a query),
- the data itself (the object of a query),
- the process of accessing the data (the predicate of a query) [2].

C. Bizer [13] derived three levels of information quality metrics in web-based information systems:

- **Content-Based Metrics** use information to be assessed itself as quality indicator. The methods analyze information itself or compare information with related information.
- **Context-Based Metrics** employ meta-information about the information content and the circumstances in which information was created, e.g., who said what and when, as quality indicator.
- **Rating-Based Metrics** rely on explicit ratings about information itself, information sources, or information providers. Ratings may originate from the information consumer herself, other information consumers, or domain experts.

We adjusted these three levels to assess iq for qKAI needs to **first, second and third level assessment** divided into **Metadata analysis, user interaction** and **intelligent analysis**. There is no absolute quality, but we can compare resources with each other (Open World Assumption) and weight them based on the amount and structure of metadata, for example. Enrichment of a resource happens in a corresponding qKAI URI by semantic interlinking and annotation. Ranking according to available metadata properties or interaction history is possible too.

A. First level assessment: Metadata analysis

According to Bizer this level enables **Context-based assessment** of metadata directly related to a resource like format, timeliness, author, provenance or language, which can be automatically detected. Metadata can be seen as a quality feature. The more metadata we are extracting, the better we get to know the content. In qKAI we are implementing the support of Aperture [20] to fetch e.g., Dublin core elements [21] like listed in **Table 2**.

TABLE II. EXEMPLARY DUBLIN CORE ELEMT SET FOR METADATA [21]

Element	Definition and recommended value formats
Title	A name given to the resource. Value format: Free text.
Creator	An entity primarily responsible for creating the content of the resource. Value format: Name as free text.
Subject	A topic of the content of the resource. Value formats: Library of Congress Subject Headings (LCSH), Medical Subject Headings (MeSH), Dewy Decimal Classification (DDC).
Description	An account of the content of the resource. Value format: Free text.
Publisher	An entity responsible for making the resource available. Value format: Name as free text.
Contributor	An entity responsible for making contributions to the content of the resource. Value format: Name as free text.
Date	The date when the resource was created or made available. Value Format: W3C-DTF.
Type	The nature or genre of the content of the resource. Value Format: DCMI Type Vocabulary.
Format	The physical or digital manifestation of the resource. Value Format: MIME-Type.
…	…

Comparable iq scores can be derived out of adjustable quality policies like e.g., available metadata property count: The less metadata properties a resource contains, the smaller is its iq score in believability or reputation. Even provenance and timeliness are very important aspects concerning trust in a resources' content. Information about the author is also very relevant for the resources quality. A user with high personal scores in certain knowledge domains has high reputation in this area. We can speak of local reputation here, because it is dependent the same way, the iq-criteria are, from task, user and context.

B. Second level assessment: User interaction

We allocate criteria here that can be assessed with the help of user interaction. Questionnaires are often used to get feedback from the user for this purpose. According to Bizer this is called **Rating-based assessment.**

The user can help e.g., to enhance accuracy even regarding semantically correctness. To evaluate factual knowledge like "*Berlin lies at the Spree*" or "*Hanover is the capital of Lower Saxony*", we see user rating and ranking following the established Web 2.0 manner as an effective solution to mark wrong content and to rank valuable or popular content step by step. Next to this crowd sourcing community approach we offer role- and level-based quality control mechanisms. Lecturers earn rewards while rating and creating educational resources out of Open Content; students earn rewards while answering questions, managing gaming tasks, exploring further content or ranking their favorites. Step-wise content can be qualified this way. Resources are marked following their quality level as **reviewed, proofed**

or not yet qualified to enable embedding in different levels of knowledge transfer and learning. Integrating online assessment components like multiple-choice or assignment question types into social oriented software seems to be a new approach – as far as we know. Although, online assessment and rating mechanisms have many things in common and can be complementary, their combination is not mentioned so far.

C. *Third level assessment: Intelligent analysis*

By **Content-based assessment** employing Natural Language Processing to detect some more information hidden inside a resource. Aperture [20] and Virtuoso Spongers [22], for example, enable comprehensive solutions for these tasks. In case if more text engineering is needed, there are comprehensive solutions for standard Natural Language Processing (NLP) tasks (e.g., by OpenNLP [23]) to perform sentence detection, NER (Named Entity Recognition), POS (Part-Of-Speech) tagging or even semantic chunking. **Table 1.** shows the related iq-criteria from relevant literature. If we talk about information quality, we also talk about user preferences and personalization. It is obvious that many of the iq-criteria are relevant while user interaction takes place, because they are subjective – user, task and context dependent. Most of the iq-criteria have direct impact on the users' interaction. There are only a few iq-criteria like "*amount of data*" or "*completeness*" that can be assessed with little or no user interaction at all. Even technical criteria influence usability, ease of use and user motivation elementary. Without fulfilling e.g., technical criteria in a sufficient way, smart interaction is not possible at the user side. Altogether, the 2nd level of our qualifying model with strong focus on user interaction is the most important and influential one if we want to determine relevant, but subjective iq scores.

V. IQ ASSESSMENT WITH THE HELP OF ENHANCED USER INTERACTION

Incentive for user participation is implemented as globally rewarding system of any interaction in qKAI (qPOINT, qRANK). **Table 3** shows interaction types, their assigned reward in form of gaming points and improvable iq-criteria. Every interaction is based on a resource. We are implementing different types of interaction like described in the following.

TABLE III. INTERACTION TASKS, ASSIGNED REWARDING POINTS AND IMPROVABLE IQ-CRITERIA

Interaction	Reward	Improvable iq-criteria
Edit	+50 points	Accuracy, consistency, objectivity, timeliness, believability, reputation, completeness, understandability
Create	+100 points	Completeness, accuracy, verifiability, amount of data
Annotate/ add/interlink	+50 points	Completeness, accuracy, verifiability, amount of data, interpretability, understandability
Rate/rank	+10 points	Relevancy, accuracy, believability, reputation, objectivity, interpretability, understandability, rep. conciseness

A. *Simple and direct feedback*

Like in common surveys and evaluation, rating happens by questionnaires with predefined scores. These ratings can evaluate persons, resources or knowledge units.

B. *Enhanced feedback and game-based interaction*

Every resource that is visualized or just queried by qKAI can be rated and ranked by user interaction or automated metrics like metadata detection. The more a resource is requested, the more statitistically data we gain. The more we know about a resource, the better we can personalize its usage.

Next to edit, create, annotate, add, interlink and rate resources and users we offer the following game-based options. qKAI jokers allow game-based functionality to add additional sources and to qualify metadata by rating and ranking input to the qKAI knowledge base. Playing the "*Know-it-all-Joker*" bounds the user to add a source (or information) that proves contrary statements. The "*Nonsense-Joker*" marks an information unit as semantically wrong or inconsistent and defers it to review mode by other qKAI users. The "*Hint-Joker*" allows looking up related sources or other users' answers as solution suggestion. The "*Explorer-Joker*" allows exploring the right answer on the web outside of qKAI during a predefined time. The "*History-Joker*" enables lookups in played answers, ratings of other users by logged interaction and transaction protocols. Statistical protocol analysis is suitable to infer further metadata.

C. *Indirect and automated feedback*

History protocols and interaction recording allows to deduce statistically results for rating and ranking purpose. Therefore, Simple Scoring Functions, Collaborative Filtering, Web-of Trust algorithms or Flow Models can be deployed in the future.

VI. THE RELEVANCE OF PICTURES IN FOLKSONOMIES

In this section we introduce one of our example use cases to enhance and determine the quality of Open Content by **collective intelligence**. Tagging is very popular in online communities these days. Everybody can participate in tagging content. Tags offer a wide range of keywords but are subjective as well and might be confusing sometimes.

A. *Relevance of pictures*

Focus here is the quality of the images found on the web. With Flickr [40] a highly demanding data source with more than two billion images and over two million new images per day is given. A crucial problem that has emerged during the study was the relevance of the found images. Many images that are found with the help of the Flickr web service do not clearly correspond to the search term. They do not deliver the desired content. The challenge that arises from this is the automatic sorting of images according to their relevance. Flickr offers a very comprehensive interface (API) which allows more possibilities than the pure web service. We developed a small application called Flickr-analyzer that is used for analytical purposes. The search for clean images, in contrast to text or text-picture combinations is difficult because they contain too little information to be found. [24]

"*There are many resources which are not searchable in folksonomies because they do not contain most of the relevant tags*" [25].

One way to facilitate the search of images on the Web is additional metadata e.g., by adding tags of our own choice. This kind of annotation is different from professional annotations in that they do not use notations and relations. Basically annotations facilitate the search and navigation of resources. The common form of this annotation is referred to in the latest generation of the Web as **collaborative tagging**. Services that allow this type of metadata generation are known as **tagging systems**. The most famous among them are Flickr [40], YouTube [44] and Del.icio.us [45].

B. Tagging and tagging systems

If user index resources with additional keywords called tags, this is called "*tagging*". There are two types of tags: normal tags and machine tags. The former are from users randomly selected keywords that reflect mostly the image content or additional information about the resources. Machine tags are machine-generated tags. These include auto-tagging and tags, which have a certain shape. Geo-tags are information indicating the geographical coordinates of the origin of the pictures or the coordinates of objects, which are shown in the pictures. Web 2.0 services that allow collaborative tagging are known as tagging systems [26]. Tagging not only organizes the resources in tagging systems in a better way, but also means that a collaborative network is formed.

"*Social tagging is used by users to build both its own network, as well as the network to" watch ", and get as new sources for the topic areas of interest*" [27]

C. Geo-Tagging

Geo-tagging is composed of two words, "*Geo*" and "*tagging*" and describes the geographic positioning of e.g., images. Many images are from a GPS receiver located at the camera automatically. This means images will be automatically marked with longitude and latitude. In Flickr, users geo-tagged their photos on a specific format: geo: lon = 13.127787 geo: lat = 52.393684. This allows an image to be found with the help of the coordinates. In Flickr, people upload over three million geo-tagged images per month [40]. For the organization and search of resources in tagging systems tags are a very important source of information. The quality of the image search is highly dependent on how well the image with keywords, called tags, is annotated [28]. A visual representation of the vocabulary used in these tagging systems is known as tag clouds. To gain a better

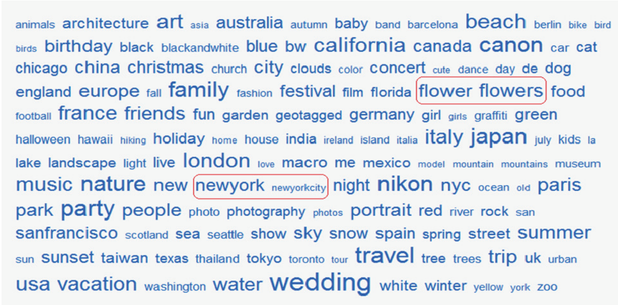

Figure 2. Flickr image tag cloud

understanding of the use of tags to obtain, the following will present the so-called tag-space and the associated tagging behaviors of users are examined more closely. The analysis refers mainly to the photo community Flickr and the bookmarking service Del.icio.us. **Figure 2** represents the most popular tags from Flickr in a tag cloud. This type of representation is an alternative to the classical search by text. It allows that users access also information that they have not sought explicitly. They click their way through the tags to images or to others that are similar. The tags in a tag cloud will appear sorted alphabetically. The size of the font depends on the frequency of the tags. Not too surprising is that the terms are chosen very general, since only these are used by most users. Striking here is mainly that some of the keywords differ only in the singular and plural (flower, flowers, or girl, girls) or abbreviations of another (and nyc newyorkcity). To express it only in numbers: there are about 5.5 million Flickr photos tagged with "*nyc*" and about 7.5 million other images are annotated with "*New York*". This means that many images actually reflect the same context, but are not found because they were not indexed consistently. A user writes in a Flickr discussion forum:

"*Is there anything Flickr admins can do about people not tagging their photographs with relevant tags. I'm tired of finding random naked people when searching for baseball shots* " [40]

This raises the question: Are user really tagging in the common interest? The response of another user on it:

"*Tags are for the people applying them, so, although they may have no relevance to you, they may have relevance to the person tagging*" [40]

Users annotate their resources primarily of self-interest. Terms they use may be relevant for them, but in the common interest they are rather irrelevant. An added value to the community arises primarily, if users annotate photos from other users, as they choose in this case rather more objective descriptions.

VII. ANALYZIS OF TAGGING SYSTEMS

The fact that social tagging is free of ontologies makes it simple for general use but more difficult for machine evaluation. A good classification (taxonomy) is essential for a large amount of data. The use of a tag of more than one person can provide a common classification scheme. Tags can be recognized as a connection between users and resources. Which users share a tag and what resources were annotated with similar tags is important analytical information in research with folksonomies.

A. Folksonomies

Users can annotate resources in a tagging system. In the literature this is referred to as collaborative tagging. The collection of tags, created this way is called **folksonomy**. The term "*folksonomy*" consists of the words "*folk*" and "*taxonomy*" and is attributed to Thomas Vander. Taxonomies are classification systems for data, which are usually hierarchical. Unlike taxonomies, folksonomies have no hierarchical structure and are not developed to purposes of classification, but arise automatically as users tag resources. The advantage of folksonomies is their simplicity, since users have complete freedom in the allocation of tags. There are two types of folksonomies: broad and narrow folksonomies, which is crucial for the analysis of tagging systems.

Broad folksomony

In broad folksonomies, many different users (user A to F in **Figure 3**) an index of content creators is made available to any document or similar tags. Thus, the document content from a variety of different or the same subject headings is described [29] . Users that use the same keyword are grouped together in groups and find the document on the basis of this keyword again (arrow in both directions). The most popular Web 2.0 application that uses this kind of multiple annotations is the social bookmarking service Del.icio.us.

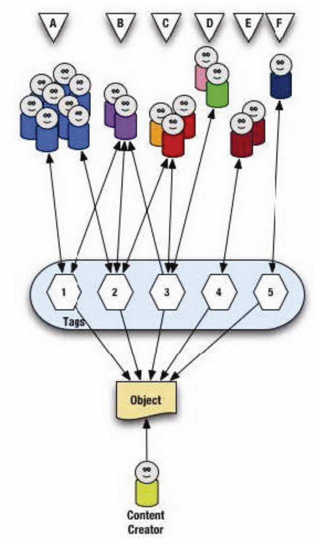

Figure 3. Broad folksonomy [29]

Narrow folksonomy

In narrow folksonomies (**Figure 4**) tags are issued only once and recorded only once. Therefore users can only assign new tags. There is no way to count and tag frequencies observed distributions. Mostly the author (or the content creator) creates the first tags. Sometimes it is also allowed to other users to add additional tags. Web 2.0 services that work with narrow folksonomies include Flickr, Technorati, YouTube, a.o..

B. Weaknesses of Folksonomies

The classification of resources by folksonomy users itself is a problem because the tags are dependable from their own view. This view is understandably subjective, and therefore needs not always to agree with other folksonomy users. This subjectivity limits the retrieval of a resource within the folksonomy. Similarly, ambiguity is problematic for the retrieval of resources because they deteriorate the precision of the keyword search. Here, the precision of the results is enhanced by reducing ambiguous terms and the yield of synonymous words, which were not included in the keyword search. This weakness could be an offset by the use of ontologies. Ontologies enable the creation of semantic relations to represent different levels of abstraction and thus express the relatedness of individual elements. At the same time ontologies allow support for synonyms, homonyms and multilingualism. Ontologies can handle the annotation of resources more efficient, as well as open up extensive search option. Synonyms can be recognized and included in the search: Who is looking for "*Brasil*", is also looking for "*Brazil*". The display of related concepts can guide the search in the right direction: If you are looking for "*mac*", you could also be interested in "*osx*". Upper and sub terms can extend and refine the search: If you are looking for '*newyorkcity*' perhaps in particular for "*central park*" or more generally for "*usa*". Recent research in folksonomies tries to analyze the importance and relationship of keywords. Most of the procedures are based on the co-occurrence of two tags. The calculated co-occurrence value is the number of resources where both tags together occur [30]. We analyzed concepts like the **Actor-Concept-Instance model** and **similarity measures** [25], [49], [50], [51] that derive ontologies out of folksonomies. For detailed information about this topics please see [24]. The insights gained from these concepts will be used in **Section VIII Keyword-oriented group search and ranking in folksonomies** to come up with our own approach for the problem of relevant image search.

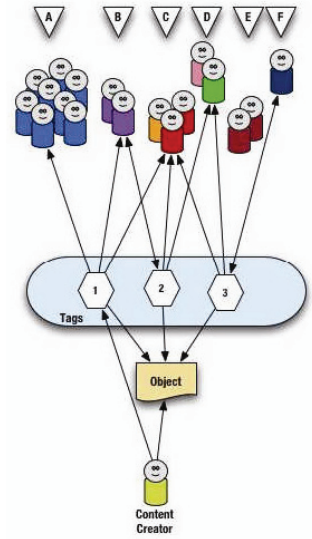

Figure 4. Narrow folksonomy [29]

C. Quality metrics for folksonomies

The absence of a single controlled vocabulary makes it difficult to assess how the quality of a tag is in relation to the retrieval of the resource. It is believed that the quality of search in tag based systems can be improved if you tag with inter-subjective meaning (a state of affairs for several viewers equally recognizable) or tags that were used by a larger group, determined automatically. This method of analysis, however, is only suitable for systems in which a term may be given more often. Term frequency within a resource is not allowed in narrow folksonomies. In broad folksonomies they provide important analytical information. The resulting power-law distribution of tags can be used as a basis for the analysis of broad folksonomies. You can concentrate in the search only to the so-called power tags.

"*We hope that offering power tags as a search option improves the precision of search results. We can justify this assumption by the opposing relationship between recall and precision. The one rises, the other falls. In the case of the search only after power tags, the recall - because the entire document-specific "long tail cut off" - is drastically reduced* [29].

In this work term frequency is used for the ranking in folksonomies (see Section C. Flickr groups). In [31] term frequency for the selection of relevant terms are used. Three metrics for the automatic selection of inter-subjective tags are presented for broad folksonomies and tested on a dataset of del.ici.us:

Metric 1: frequently used tags
For each tagged resource, the tags are sorted by the number of frequency and the five with the largest occurrence are elected. If a term has been used by several users for a particular resource, this term for an objective description is more relevant.

Metric 2: tag congruence
A tag consistency for resource x is defined by the tags that were selected by at least half of the users. This value can be achieved by dividing the number of all different tags for a resource with the number of users. Decisions in various areas of human activities are often made on the basis of the majority. More than half of the people fit in the rule of the majority and often use terms that were already in use. In broad folksonomies tagging may be like a vote for the semantic labeling of a resource [31].

Metric 3: TF-IRF weighting
For each tag the Term Frequency Inverse Resource Frequency (TF-weight calculated IRF) is calculated and only the tags by the five highest values are selected. The TF-IRF metric is derived from the term frequency inverse document frequency (TF-IDF). TF-IDF is a standard measure in the field of automatic indexation, to find the best descriptions for documents. When choosing a tag for a particular resource, the TF-IRF formula is taking into account the frequency of keywords for the document. The higher the TF-IDF value, the more valuable is the concept. For the calculation of the TF-IRF value a corpus of similar resources is requires. You get this on by clustering [31] with the Markov Clustering (MCL) algorithm that creates a graph from the co-occurrence of tag pairs. The TF-IRF value can be obtained with the TF-IDF formula conversion [31].

The three metrics were presented to a record of del.ici.us tested with 3.4 million users from different bookmarks from 30,000 in 2007. Then the test with an online survey was bound to find the most appropriate metric. **Metric 1** (**frequently used tags)** provided the best results [31].

Evaluation of the approaches
Most of the described approaches and ideas mainly work with co-occurrence, simple clustering algorithms or the vector space models. The resulting similarity values can serve as a basis for a similarity graph. In the Actor-Concept-Instance model resources, users and tags are represented as nodes. For the relationships of the tags are only the connection graph "*user tag*" and "*tag resource*" decisive. The former provides an ontology based on users with similar tagging behavior and the latter an ontology annotated on similar objects. This type of graphical modeling of tagging systems is an important basis for ranking systems, such as the **FolkRank** [32]. The algorithm is based on the idea of the PageRank algorithm and is used for the ranking in folksonomies. The PageRank algorithm computes rankings on node with the idea that a node is important if many other important nodes point to this node. Based on the FolkRank algorithm, this means that a resource is then important if it is connected with important users or tags. The FolkRank is a modification of the PageRank algorithm, as this cannot be applied directly on folksonomies. The FolkRank algorithm determines a lot of relevant resources and users for a tag. This information can be used to assist the user in the annotation and in the search.

The view of each user on a resource is subjective. Many resources (images) are ambiguous and are therefore interpreted differently by different users. The degree of content development is crucial. Some users use more general terms such as "*animal*", while others are more specific such as "*dog*" or "*puppy*" that complicate the search of resources. Users can also describe the same or very similar pictures with different keywords. While a user an image with "*lake*" annotated, this may be another tag with "*sea*". This problem is to use the surrounding to identify tags that are based on co-occurrence relationships. The **co-occurrence relationship** is highly dependent on the amount of data. For a very large amount of data (like Flickr), it is relative, since one in very many different resources for two very similar tags like "*animals*" and "*animal*" can have a low similarity value of 0.06. The main reason is that usually the tags are assigned mainly to Flickr only by the creator and he did not worry about the plural, singular or synonyms. An image that was tagged as "*car*" will probably not be additionally annotated with "*automobile*" by the same user. There are 10 times more images in Flickr tagged with "*car*" rather than "*automobile*". Procedures that try to build a threshold value from the tags top and narrower relations have the problem that many special tags are collected as generalized tags. Therefore, these are suitable only for supporting the user in his choice of auto tags and less for an annotation. Another important factor is the multilingualism. Members use many resources for different languages. Usually the mother tongue is combined with English. In addition, users can annotate a resource from different backgrounds together what is an advantage for the general search, but it brings considerable problems for machine evaluation. Many tags mean the same thing but because of different languages they have a small co-occurrence and reduce the effect of similarity calculation further. It is difficult to reach a clear classification of the tags solely on the information of the co-occurrence frequency and the frequency of tags in a library. The co-occurrence frequency allows that the less descriptive tags (which are rarely used) are eliminated.

The approach to combine folksonomies with existing ontologies provides lightweight ontologies. It filters the irrelevant tags and finds relationships between relevant concepts. The problem of ambiguity can be minimized over the Semantic Web ontologies. Considering the enormous amount of data which e.g., Flickr provides (over 2 million images per day), this is too complicated, but for limited amounts of data very demanding. The approach adopted here identifies the relationship between tag pairs on the semantic search engine Swoogle that has only the English language. Because tags are often multilingual, this approach is suitable for images that are tagged in English only, and is less effective for multi-language terms. This problem could be limited, if we automatically translate any foreign tag into English. Such an application is presented in [33]. It translates the search terms automatically in up to six different languages. In combination we can get multilingual image retrieval from Flickr.

Quality metrics for folksonomies are suitable for the selection of relevant tags very well. Unfortunately, these mainly take into account the term frequency applicable only in broad folksonomies. Narrow folksonomies cannot show certain frequency distributions of tags since all tags are equal (all tags come only once). Therefore, the presented metrics work only for broad folksonomies. A direct application to narrow folksonomies does not provide the desired effect.

The indirect concept is the **gradation of the tags** within a tag list – so we can develop other methods to determine the **relevance of resources' tags**. One solution for this is presented in Section VIII D.

The relevance of the assigned tags is critical for the retrieval of the images. We presented some approaches that examine the relevance of keywords. Since the quality of tags is dependent on the co-occurrence relationship and therefore on the tagging people the similarity graph is an efficient modeling method for folksonomies. This helps to consider the tagging behavior of users, the co-occurrence and term frequency simultaneously. Unfortunately, this information alone is not enough to improve the quality (relevance) of the tags automatically. But the information is well suited to support systems in proposing tags to the user. Some approaches attempt to get additional help by external sources such as Wordnet , Wikipedia, Google or the Semantic Web search engine Swoogle. These make it possible to find a genuine search for synonyms or discovered ontologies. Synonyms help eliminate the significance of ambiguous tags.

In the next Section we introduce our own derived ideas and approaches to allow **image search optimization in folksonomies**. For experimental purposes only we use the Flickr online photo community. Flickr provides next to the API and the tags other metadata such as description of images, comments and number of clicks (views). This information can be used to make a statement about the quality of the found images.

VIII. Keyword-oriented Group Search and Ranking in Folksonomies

Groups allow pre-selected content and increase the precision and relevance of the recall. Our idea to improve search results is a keyword-oriented group search and ranking. We developed a tag ranking game called qRANK to rate and rank Web resources. Flickr allows its users to organize pictures in groups and related groups in collections. Groups, tags, views and comments are important information to learn from folksonomies. The aim of this work is not to develop a global algorithm for the complex search problem in folksonomies. Rather, we implemented and evaluated ideas and methods to optimize photo relevance and quality for Web photo searches. A methodology which allows an automatic classification and ranking of photos of their attractiveness was developed in [35]. Photo attractiveness is a very subjective term that depends on many factors. The feedback from the user will supply important information for classification and regression models to create, based on visual characteristics of images and the metadata

„*In a wider system context, such techniques can be useful to enhance ranking functions for photo search, and, more generally, to complement mining and retrieval methods based on text, other meta data and social dimensions.*“ [35]

Visual features such as "*color*", "*contrast*" and "*rudeness*" of images and other metadata such as tags and favorites lists are examined. The combination of visual and textual features yielded the best results for the ranking according to a photo's attractiveness.

Here the main issue is the quality of the image search. The quality of a search result is determined by the intention of the searcher. Therefore, it is an advantage to consider the search behavior and motivation of the user precisely. In general, a user has the following interests:

1. **Precise search**: the user is looking for a specific image or images for example of the Eiffel Tower.
2. **Search topics**: he is looking for a picture or pictures on a specific topic such as only black cats or dogs of a particular race.
3. He has **no particular intention** of looking rather out of curiosity and wants a closer look at village (vicinity search).

A. Attractiveness of pictures

This approach should help to determine the precision of the images by the attractiveness and popularity of the photos. A scenario for an exact search might look like this: A user searches for a picture of the new city hall in Hanover to use

Figure 5. Flickr standard search for the terms „Rathaus“ and „Hannover“

in his school lecture. He used the two keywords "*Rathaus*", and "*Hannover*". Therefore the **standard keyword-search in Flickr** provides 175 results. We can display the first ten images at random and get the following pictures as seen in **Figure 5**. Also there are some images on the town hall, none of this is what he really wants to use for his work. Of course, among the 175 photos found there are some that correspond to his ideas and with a little patience he would find the right image. However, the user wants to find the photo that is relevant to his search as soon as possible. The relevance of the image here refers to the given information content for the user, as generally all images may be relevant. The intent of the user (use: seminar work) implies that the content of the image must satisfy the search term clearly. Relevance is indeed a relationship between an image and a user. A tag and a picture are defined as relevant, if the tag only describes aspects of the visual content of an image [36]. In the course of this work we call **relevance** (also used in precision) the degree to which the content of an image corresponds to the entered search criteria. This degree of precision can be used to classify images. Besides the problem that many images cannot be found because they were annotated with little or inaccurate tags, there is a further problem, to assess the degree of relevance. For some queries you get a very large selection of Flickr images that are different relevant. Since one is usually interested only up to a fraction of these images, a ranking of the found images is required. There is a patent publication of Yahoo! for Flickr which deals with this problem [29]. There are set five criteria for a ranking by interestingness in narrow folksonomies:

1. The number of tags to a document
2. The number of people tagging a document
3. The number of users that get the document after search
4. The relevance of the tags
5. The time (the older the document, the less relevant)

Most of these criteria are closely related. The first two criteria are important for the relevance of the tags. If multiple users annotate an image with different terms, they create a multidimensional view upon the resource. Suitably chosen tags facilitate the search. If the terms are very different, the search is inaccurate. An image that was tagged by different users reflects also the popularity of this picture again. Photos which are described with many tags are found more often. The criterion of time is not applicable, because a picture does not lose its relevance over time. The feature "*Interestingness*" is described in Flickr [40] as follows:

"*Many factors affect whether something is on Flickr interesting (or not). It depends on the origin of the clicks, who commented when the image of who identifies it as a favorite, which tags are used, and many more factors that change constantly*."

As the components are related is deliberately not discussed deeply. Derived from [29] we define three different sets of criteria for the ranking in tagged documents (see **Figure 6**) which are of importance for our work.

The first volume contains procedures that relate to the semantics of the tags. The relevance of the tags can be determined using the method presented in the previous section as the TF-IDF weighting, the cosine similarity or the FolkRank algorithm. In addition to these criteria, there are other factors, such as click-through rates, the number of comments and favorites list, which can be crucial to a relevant search (collaboration). In addition, you can include the relevance of terms, the feedback of the users with (prosumer). This can be done in a question-answer game where users assess metadata of resources playfully.

For a relevant search, some of the investigated options shown in **Figure 6** are examined. In the next approach, we use the **click-through rates** and the **upload date** of the pictures and would like to examine whether images, which are often viewed at the same time have a higher relevance. About the interface of the Flickr API can about each picture about click rates (views) and the upload date to be fetched. The number of clicks is an implicit relevance feedback, "*they are in a high degree collaboration-oriented ranking criterion in the sense of Web 2.0*" [29]. The mark as a favorite reflects the attraction and popularity of the image. In general, one can assume that with increasing click rate, the favorite rate rises. Thus, we extended our search with an additional function that sorts the pictures by clicking the spending rate. The click rate is a picture of the dependent "*upload date*" dependent. Photos that are longer online have generally a

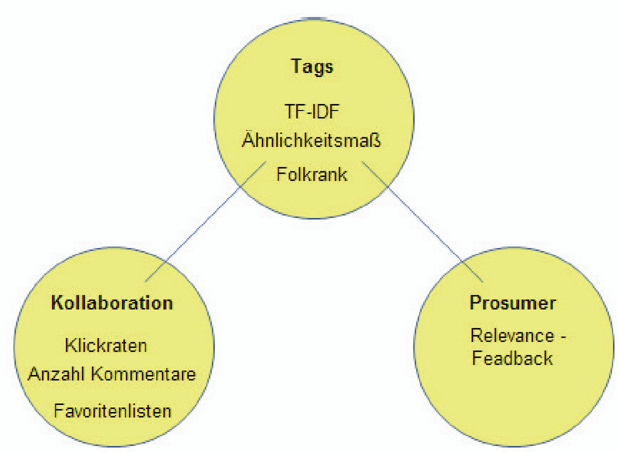

Figure 6. Ranking criteria in folksonomies

Figure 7. Extended Flickr search for the terms „*Rathaus Hannover*" with precision formula

higher click rate than actual pictures. To counteract this, the upload time in the calculation considered. This function is called the **precision formula**, resulting from the division of the click rate and the time (in seconds) that a picture is already online sets together. About combines the precision of the ranking formula for "interestingness", the relevance of the retrieval set is clearly improved. The same search from the previous example, sorted according to the precision value, returns the data shown in **Figure 7** with the first ten images that the user receives after a search for "*Rathaus Hannover*". The weakness of this method is that the images are very new, get assigned a higher weight than older ones. An image that has ten clicks on the first day would have a very high precision value without being necessarily relevant for our search. The click through rate alone is not an absolute indicator of the relevance of a search. The click-through rate of an image rather reflects their popularity again. This in turn depends on several factors. As a rule, to Flickr photos that belong to a broad community often looked at. Images that contain many groups, and its creator are linked with many other users have generally higher click rates. This means that the pictures were annotated and rather inappropriate for a subject search are not relevant, but can have a very high popularity. In Section VIII C., an approach is presented, how images have grown to their relevance.

A major problem in the search for relevant images is the ambiguity of the tags. The tag "*Paris*" can mean a city in France or a city in the U.S. or even refer to a name. When the user searches the tag "*Paris*" for pictures of the French capital, he will receive, among other things pictures from America or from people who are called Paris. This can reduce it but if we expanded the query with related terms. In the research of folksonomies this approach is the "*tag of suggestion*" [37] or called tag recommendation [28] [30] and can be used for two things. First, you can use it to help the users to support the annotation. Recommendations will help users to clarify the image content as well as reminding them of related semantics which may otherwise be ignored [28]. On the other hand we can extend the inquiry with other tags in order to achieve a more relevant search. We concentrate here on the second.

B. *Tag suggestion*

The idea of tag suggestion is used in this section to specify the search for images. From previous considerations we know that tags are ambiguous, imprecise and often irrelevant. Linguistic differences and the fact that users are not professional tagger make it difficult to find the pictures in Flickr. If a user has annotated a picture with the words "*cat*", "*white*" and "*charly*" we will not find this picture, if we search for the keyword "*Katze*" (German translation). In Flickr, there are twice as many images that are tagged as "*cat*" than with "*Katze*" and also about the same as many pictures that are tagged with "*cats*" instead of "*cat*". Even if these images actually reflect the same content, they form different result sets in Flickr. Some works in the tag list folksonomies combine an image with relevant concepts from other sources such as WordNet [36]. In this paper, we focus primarily on the query and try to isolate the problem of imprecise tagging, as we show related tags to the user automatically. Here the question is expanded by the user with the selected terms. Based on the above example, the user gets a list of related tags containing terms like "*cat*" and "*cats*" while searching for "*Katze*". These are terms that often occur together with the search word (co-occurrence relationships). On extending the search to several terms, also increases the amount of results.

The query extension can be used to further narrow down the search space. This is e.g., in qMAP used to reduce the problem of synonyms. If a user searches for the word "*apple*" searches, it is not clear whether this term refers to the fruit "*apple*" or to the company "*Apple*". Such an inquiry would yield many irrelevant images. However, if the request is extended with an additional term such as "*fruit*" or "*Mac*", then its ambiguity is eliminated. In this simple case, the searcher possibly finds out on his own that his request is not clear and would change or expand his search with a further term. In most cases, however, a user does not worry about whether his chosen search term is ambiguous and much less he finds an appropriate term with which he can formulate his question precisely. An improperly selected tag means that the results are again irrelevant or relevant images are not found. A selection of tags that are related to the term used by the user in a strong correlation facilitates the search. In qMAP, the user gets a list of related tags available for selection like in the query extension. The terms selected by the user are involved in the request and only images are displayed that contain the tag list and all of the keywords. A multi-query search is also suitable for general subject searches: A user searches for a specific topic such as black cats. This is the request for "*cat*" extended with the term "*black*" and searched for images that contain both words. In response, the user gets only pictures that at least contain the two concepts "*cat*" and "*black*". For a more precise topic search, this version is less suitable. From the knowledge that many images are annotated inaccurate, it can be assumed that the method of query expansion also provides images that do not contain any black cats. On the other hand, there are also pictures that would have been useful to the user on the context, but are not found due to the lack of tags. The number of tags per image is very limited in Flickr [40]. This is because most of the pictures are annotated only by the creator and are not tagged with many words. In addition, a user does not take the time to worry about and to discuss alternative and more detailed tags. In contrast, the **groups at Flickr** are used more often. A study in [38] has found out that over the half of the users (about 8 million) share at least one Flickr photo with a group. Flickr groups are self-organized communities with common interests [38]. The analysis of Flickr groups is an important step to find relevant images that were inaccurate or not tagged. In this study, the groups are used primarily for the **subject search**.

C. *Flickr groups*

A group is a collection of people and objects that are either in physical proximity or share certain abstract properties. The main goal of a group is to facilitate the exchange of resources in a community. In contrast to the similarity graph in previous sections, groups are not generated algorithmically. They arise spontaneously, not by chance:

"*Users participate in groups by sharing and commenting on photos, most often on specific topics or themes, like a popular event, location, or photographic style.*“ [38]

Such collective behavior modes offer alternative ways to understand and analyze visual content. Grouping is a simple and well-received folksonomy function, which provides valuable information to detect relevant resources and improves the quality of the search [41]. Most groups had a clear theme, and are sorted in this context issues.

"*Two images are similar if they belong to the same Flickr group*" [47].

Users who are involved usually have the same interests. They exchange information and knowledge by group discussions and comments about the pictures. The resulting **collective intelligence** enables that the images are better annotated in well moderated groups. Members, who are friends with each other, develop similar approaches to an image. In [42] the effect of the grouping in a tagging system is presented with **Group Me!,** in which the user can organize any resources from other tagging systems in groups via drag-and-drop. Group Me! allows not only tagging of resources but also tagging of the groups themselves. The annotation of resources can always be considered in the context of a particular group. This provides additional relationships that can be used for the quality of the resource ranking:

"*Tagging resources is always done in context of a certain group. This group context gains new relations between entities of the GroupMe! folksonomy, which consists of user-tag-resource-group bindings, e.g., the group's tags are likely to be relevant for the members of the group, and vice versa. Such new relations enable advanced folksonomy-based ranking strategy.*" [43]

A ranking algorithm is in Group Me! presented that uses the effect of the grouping for the ranking in folksonomies. The “*Grank*” algorithm based on FolkRank returns through the use of the group structure better results than the general FolkRank algorithm [43].

In Flickr, groups are collections of people who join voluntarily in a community. The collections of resources that are collected by the group members are called “*group pool*”. Each user can create any number of groups. There are three different types of groups that are crucial to the search for these:

(1) public, everyone can see the group photos and join the group.
(2) public, everyone can see the pictures, membership by invitation only.
(3) private, no one can find the group, membership by invitation only. here consider only publicly accessible groups.

Here, we concentrate on public groups only. In [38], the group structure of Flickr is analyzed. The average number of members per group is approximately 317 (**Figure 8**).

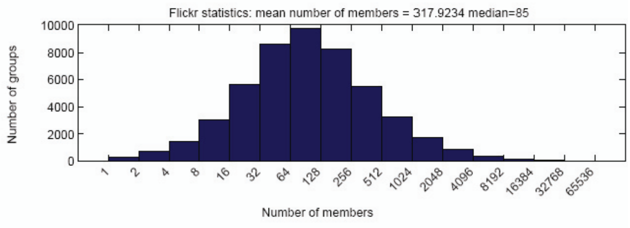

Figure 8. Analysis Flickr groups "number of members" [38]

Unfortunately, there are also many groups in Flickr with very few members and even groups without images. These provide no information in this work and are known as "*spam groups*". The average number of photos in a group is

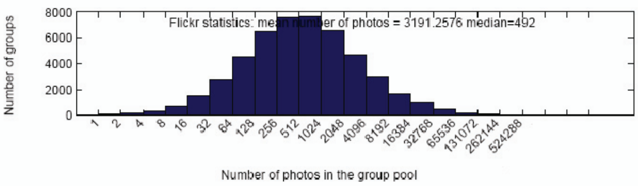

Figure 9. Analysis Flickr groups "total images" [38]

approximately 3191 photos (**Figure 9**). Both images are a proof that the **exchange of photos in groups** is an **important activity among Flickr users**. More than 50% of the users share at least one picture with a group. Over 25% of the members share at least 50 images [38]. A photo can also be included in several groups. Groups ensure a higher exposure of the photos. They offer the user a wide selection of relevant images for a specific topic and make the photos easier to find. Similar difficult to the search for images is the search for relevant groups:

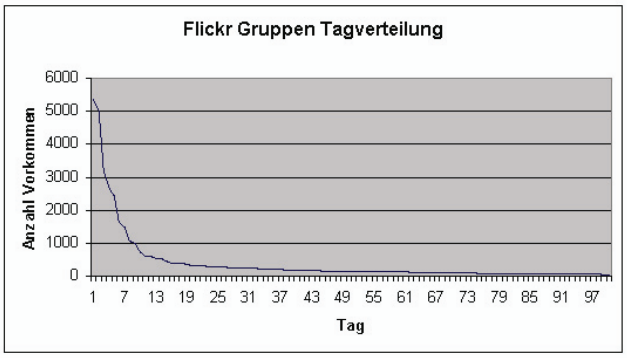

Figure 10. Example tag distribution in a Flickr group

"*In practice, finding groups on Flickr is relatively cumbersome and does not make use of the plethora of meta-data available in the user groups and photo collections*" [38].

Groups are found in Flickr in the first place through their group name or description. The title of a group is not always perfectly. The description is often too broad and not specific enough and we find irrelevant and too many groups for a specific topic. According to [38] 60% of the groups consist out of one to five relevant subjects and only in 10% of the groups we find more than ten subjects. Unlike in Group Me!, users can annotate only the pictures in Flickr. The number of tags in a group is therefore limited by the maximum of 75 words the images can be described with. **Figure 10** shows the 100 most used tags in a group with a total of 15.222 elements. At the beginning of the curve a few tags are placed with high values, the right end is composed of many nearly equivalent tags. This type of distribution that is similar to a **power law curve**, was discovered in broad folksonomies by Thomas Vander Wal [29].

The tag distribution of the Flickr groups is almost identical with the ideal power law function. The green area in **Figure 11** contains tags that are found in most resources. These reflect the collective opinion of the group members and are more relevant for the groups' subject. In the yellow area, includes the so-called "*Long Tail*" as the special tags. These are rather subjective tags that are related less to the subject in the group. There are no annotated Flickr groups, so one can derive the tags of the images to the groups when considering the groups as one resource. The tags, which

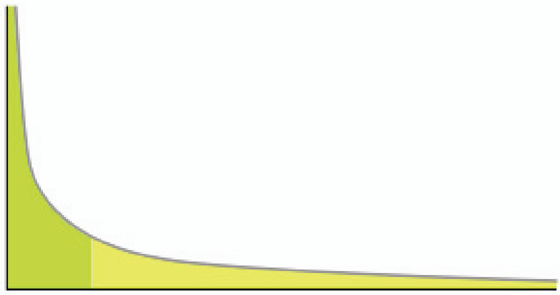

Figure 11. Power-Law curve [29]

occur frequently, are more relevant to the topic in the group. For further and detailed information about our group mechanics please see [24].

From the previous considerations we now deduce our **tag-based search and ranking procedure for Flickr groups**. The approach builds on the search methods used in Flickr, but then considers ranking of the search results by the most used tags in each group. In addition, this method eliminates groups that have little or no elements. For the ranking of the groups following information should be considered:

1.) The members and the number of elements.
2.) The most used tags with a weighting factor.
3.) The titles and the descriptions of the groups.

The idea is that groups that contain most of the pictures in the ratio for the given tag are most relevant for a subject search. Since the groups are primarily used to get the most images on a specific subject, only groups are interesting, that provide a certain amount of images. Therefore, the group ranking process ranks the groups according to the quantity of images that are annotated with the wanted keyword. The most commonly used tags are elected as representatives of the groups.

Application flow

First, the search term is compared with the most popular tags in a group. All groups that contain the search term as tag are weighted on the frequency of their tags. If we look for groups that follow a clear theme, then the weighting is based on the number of elements with this tagged term divided by all the elements. If one is interested in the most pictures to a search term, then the occurrence of this term is used as a weighting factor. If we got the group with the most appropriate images, we can do a keyword search within this group and for example sort the images according to their relevance with qRANK (see Section IX).

Then we successively take into account the following criteria:

1. The compliance of the users' search term with the groups' tags is examined.

Since users are using a known way in their annotation usually and not all forms of a term together, the above condition is extended. A user who searches for "*church*" is also interested in pictures annotated with "*churches"* and "*Kirchen*".

1.1. An English translation of the search term is taken into account in the search
1.2. To see the similarity between the plural and singular, the Levenshtein metric is applied with a distance of two.

The Levenshtein metric can be applied easily, because we can usually consider, that terms like "*Church*" and "*cherry*"

Rang	Rang des gesuchten Begriffs innerhalb der Gruppe	Mitgl.	Bilder	Top-Fünf-Tags	Bilder mit Kirche/ church	
1	4	24	123	münchen munich architecture kirche church	77	1
2	0	1	0	keine Tags	0	11
3	1	2	1	kirche parchim	1	10
4	0	1	51	judith thomas moe hochzeit	0	9
8	49	420	2358	jesus christianity hymn chant christ	143	8
9	135	6671	79863	italia italy anticando church roma	14598	7
47	5	1643	23796	church europe cathedral architecture kirche	11447	4
138	2	560	30354	church kirche carving austria österreich	8433	6
245	4	91	2173	church europe cathedral kirche architecture	1280	2
367	3	180	1441	gothic architecture church cathedral england	419	5
448	4	1235	10809	church architecture europe kirche cathedral	5217	3

Figure 12. Flickr group analyzes for the term "*Kirche*"

are different in more than two places. They are not in a singular-plural relationship and are not together amongst the most used tags found in a group because they represent two very different topics.

If a query matches with one of the top five tags, the affected groups are ranked according to the weighting factor. If several terms match the sum of all weights is formed. If the tag list of a group does not contain the search term or empty groups are weighted with Zero. All groups that are equally weighted are ranked according to a second criterion: the number of images. If the number of images is also equal, as third the number of members is taken into account. As a result of the procedure we get a ranked list of the groups. This method is especially effective if we seek for general subjects that provide a wide range of groups. **Figure 12** contains an example part of the list of results for the term "*church*", which provides a total of 1551 groups. Since the list in fact, very long, here is shown just a snippet. The column "*Rank*" in the table gives the position in the list that Flickr (sorted by the relevance) returns. The idea of this group ranking procedure is to **find the group with the most relevant images**. The red numbers in the table represent the rank that our presented method derived. At the first rank position, both lists are still identical, but the remaining positions differ massively. Many groups which "*Kirche*" in their top five tags are weighted stronger by Flickr than groups that use the tag "*Kirche*" not at all or very rare. The explicit consideration of the tags' plural/singular and the inclusion of the terms' English words come to significantly better results than the standard Flickr search. Since Flickr does not provide intentionally the needed data for the approach, they must first be created. At once, Flickr allows only a maximum of 500 pictures or information per request to download. In order to realize a dynamic and non-redundant storage concept, the idea of the Actor-Concept-Instance model has been implemented. For further detailed implementation details please see [24].

D. *Tag ranking*

The approach discussed in the previous section allows ranking the groups according to their relevance. Only term frequencies will be considered, which are calculated from the tags of the images. The images in the groups are not ranked yet. In the following, the idea for an image ranking game called **qRANK** is presented. It provides important information to **rank the tag list of an image automatically**. This information is then used to sort images according to their relevance.

Narrow folksonomies like Flickr, have a major disadvantage that they do not allow the frequency distribution of the indexed terms. Therefore, it is not possible to observe tags abundances and distributions within a resource. All tags come only once, so that we do not have simple methods to distinguish between relevant and irrelevant tags. A user can tag his pictures in Flickr with up to 75 keywords. In general, the tags are chosen arbitrarily.

With known methods we mentioned in **Section VIII A.** like TF-IDF weighting we could determine the relevance more precisely automatically. Here we like to introduce the **different approach** qRANK, which allows us to classify the tag list of an image in a **game-based** way. This game should investigate in how far the process of the players acquired knowledge in a dynamic ranking may change the tag lists quality. With each pass of the game improved the tag of an image that can be used for further analysis, particularly for the improvement of the search list.

IX. *qRANK*: A TAG RANKING GAME

Most of the analysis so far considered folksonomies that deal mainly with broad folksonomies. The resulting frequency distribution of tags examined is an important indicator to determine the relevance of one tag in reference to the description ability for a resource. This collective knowledge can provide a statement about the relevance of a tag. The implementation of the tag rankings (previous section) by **a game that implements the idea of the power-**

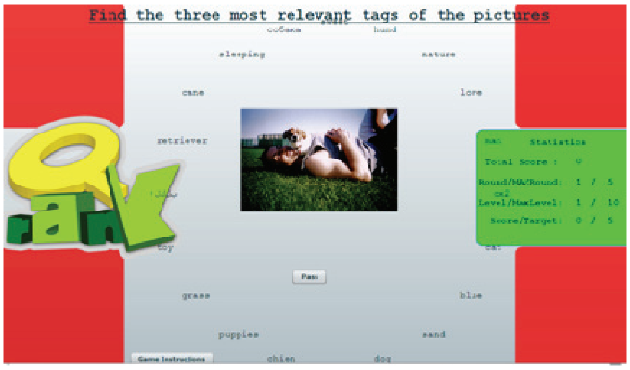

Figure 13: qRANK interface

law curve would provide **additional information** for the ranking of images. Most approaches to rank folksonomies are based much more on the FolkRank algorithm [32] or ranking techniques based on particularly elaborate calculations [33]. In this work the pictures' tag list is sorted according to the relevance of their tags. At the same time the tag list is extended and annotated with **new valuable terms**. qRANK (see screenshot **Figure 13**) queries available Web services (almost RESTful) and embeds returned content in a predefined gaming setting. Here we added some algorithms to enhance the precision (relevance) of the search results like e.g., interestingness rating or precision formulas for folksonomies. Additionally, every gaming interaction is logged and ranks played content enabling the users' collective intelligence by and by. Results are stored in qKAI but are still semantically interlinked with the provenance source not to lose the resources' context and for updating. Techniques used are semantically Linked Data (annotation, interlinking), server-side Java, Adobe Flex/Flash and a MySql database – to be flexible in representation. For further implementation details please see [24] and [48].

A. qRANK: game description

The user gets presented a picture and a list of twenty tags. His task is to choose the three most relevant tags that reflect the subject of the picture best in his opinion. Subsequently the chosen terms are reviewed by the rank in another list, and rewarded with points depending on the tags' rank position. For each term that is included among the top five tags, the player gets three points. In positions six to ten the user gets two points and for the positions 11-20 he receives one point. If the term is not included in the list or the rank is below 20, the user gets no points. The motivation of the player is to achieve the maximum number of points per round to get to the next level. The game consists of ten levels. In each level the player gets five consecutive images displayed and can reach a maximum of 45 points. The barrier from level one to two is at 20 points, and increases for each level by 5 points. So from level 6 you only come further to the next level having full points.

B. qRANK: architecture and backend

Figure 14 describes the components and the approximate sequence of qRANK. We downloaded a data set of relevant images to a certain topic from the Flickr web service and stored it in a MySQL database. The information for all the images are recorded in one table. In addition, the related tags that fit best on this subject are saved in another table. In the third table (image tag list) all tag lists of the images are managed. The image tag list consists of the terms that users have used to describe this picture in Flickr. A fourth table (ranked tag list) is filled dynamically. This is filled at the creation of the game with ten terms of the actual image and a related tag list tag. The ranked tag list contains for each term a counter, which is used to count the frequency of the term.

By chance, the player gets presented a photo and 20 matching tags. The tags will be selected for a specific principle from the tables "*related tag list*", "*image tag list*" and "*ranked tag list*". This achieves a useful combination of tags. In the very first run of a picture the length of the "*ranked tag list*" is set to twenty. While producing the amount of data every tag list will be employed with ten randomly selected tags out of the "*related tag list*" and "*image tag list*". The number of tags in Flickr images is

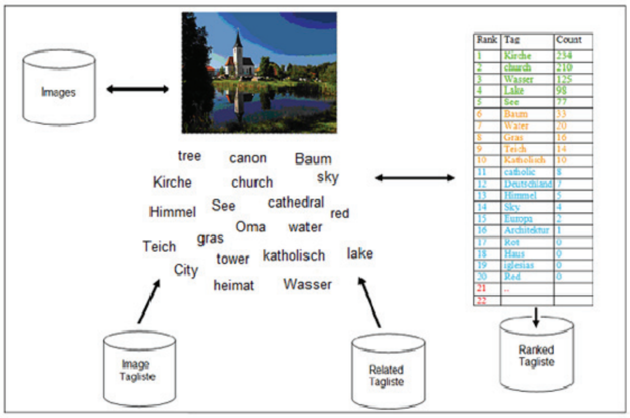

Figure 14. qRANK concept with tag lists

different; many images have less than three tags [26]. If a picture does not have ten tags, so in this case, the missing tags are added from the related tag list. These twenty tags are then stored in table "*ranked tag list*" and build the new tag list of images that is sorted dynamically through the game.

C. qRANK: gameplay

After a player has selected three terms, they are compared with the tag list and awarded with points. Since the first run of the counter of tags is to zero, an additional condition is defined: If the counter of all terms is the same, the player gets his choice irrespective of the maximum score for that round. From the second pass (for each image) is the selection tag list combined out of the first 10 tags of the ranked tag list with 5 randomly selected tags from the "*related tag list*" and the actual "*image tag list*". With the selection of the top ten ranked tags from the tag list, we ensure that terms that are more relevant are selected with a higher probability. Even here, it may happen that the actual tag list (image tag list) contains less than five tags. In this case, the remaining tags from the ranked tag list are added. To prevent duplicated tags, the randomly selected tags are compared with the related tag list and the actual tag list of the images with the first ten terms from the ranked tag list. The logic of the game is developed as web services. To get a better overview of the game's flow from the perspective of the player, it is described as follows:

1. The player gets a random image and a collection of unsorted tags. He has to choose the most relevant three terms.
2. The chosen three terms will be compared with the ranked tag list.

2.1. If they match, he gets (depending on rank of the term) points and the counter of the tag are incremented.

2.2. If the selected tag is not included in the ranked tag list, this is added thereto and the counter is set to "1". The player gets no points. This ensures that the tag list is ranked and expanded with additional terms. A limit on the maximum number of tags is not set in the game.

However, the maximum number of tags is fixed by the quantity of the tag list and the list of related terms. The primary objective of this game is to **evaluate the information gained from the existing tags to an image**. Through an extra box users can also add optional new tags.

D. qRANK: ranking of the images

The information, which is calculated from qRANK can easily be converted into a ranking of the images. Therefore, qRANK itself is already a precursor of the ranking. The more an image is played, the more meaningful is the tag list. The idea behind this ranking is similar to the group ranking. A picture is evaluated collaboratively and as a result we gain a weighted list of objective tags. The subjective tags that insignificant for information retrieval fall out automatically. Tags that do not explicitly describe the content of an image and only have a meaning for the person, who assigned them, are not included by the public (the players). The result is a tag list for each image sorted by relevance. The degree of relevance of a term for an image depends on the objective consideration of all persons who have played this picture.

The result is the basic principle of this tag ranking process. In this procedure, any tag from the ranked tag list, which belongs to the image, is weighted. The weighting consists of the simple calculation of the number of times this tag was chosen, divided by the sum of the possibilities that he stood for selection. The relevance results here out of the tag's selection counter in relation to all other tags' selection counters. A valuable statement is possible if an image is played with certain frequency.

X. EVALUATION

We have seen that the search for relevant groups and image in folksonomies represents a fundamental problem. Some related approaches have been described in this paper trying to use the resources metadata (tags to classify). From the analysis of these approaches in this work, new ideas have emerged, which were implemented as a prototype. In this section the effect of the implemented approaches in this work to search for relevant groups and pictures are shown. The experiments described below compare the **standard keyword search in Flickr** with **our group ranking method** and our **game-based approach** (qRANK).

A. Experiment 1: group ranking

The aim of the group ranking procedure is to find the group with the most relevant photos according to a topic or term. These are the groups sorted by relevance to the topic. To compare the method with the search for relevant groups in Flickr, we stored the term "*Kirche*" of 100 groups with information about the images, tags and users in a MySql database. The groups search on this term has found 1640 groups at the time of the experiment. To download all the required information over the Flickr API, we have to provide several queries for one group. Unfortunately, the Flickr API does not offer the function to determine the occurrence of a specific identifying tag at the time of this work. Therefore, an additional methodology was created to determine the frequency distribution of tags within a group. 100 groups have been considered demonstratively here, with their 100 most used tags. A data set of a million images and over 100 thousand emerged out of this. To optimize the performance of the database query the set of tags was reduced to 100 most used tags per group. The groups are selected as follows: Fifty of the groups are also the first 50, as they are returned by Flickr and the other half, randomly selected groups from the rest of the crowd.

B. Result experiment 1

Figure 15 represents the number of relevant images of the first 20 groups that Flickr [40] provides on the query "*Kirche*", compared with the process of this work. The red bars describe the results from Flickr and the green bars, the results with the group rankings from this work. During the first eight groups in Flickr together provide a total **of 100 images** to the search term, with the groups ranking procedure we get in the first position a group with **5262 images**. Considering that Flickr has all of its data available and here we included only 100 groups, the procedure becomes even more important. As we know Flickr does not explicitly take into account the tags and still less the number of images as a relevance criterion. Therefore, seven of the first eight groups in Figure 15 are empty, while the groups ranking procedure sorts the results by the number of relevant images. The relevance of the images is judged here by the strong commitment of the Flickr groups. The relevance of a group is not necessarily dependent on the number of matching images in a group. A group with fewer elements could well have more relevant images as one with more pictures. In this case, we can optimize the groups ranking method by combining it with qRANK. Thus, the tags of the images are evaluated within the groups by qRANK and the weight is derived based on the evaluation of the tags for the

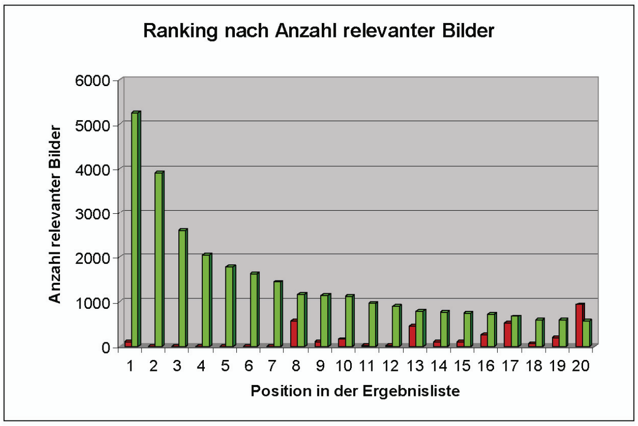

Figure 15. Results of the Flickr group rating approach

image and the group.

C. Experiment 2: game-based picture ranking with qRANK

For this experiment, we put two different versions of qRANK online for one week. The first version consisted of 250 randomly selected images to the topic "*Kirche*" and the second version of 100 images selected specifically on the topic of "*Hund*". The users should select the most relevant three terms for the image. In the first scenario a user always had to choose one of the words even if he is not sure in his choice. In the second game, the user could press a pass button to get the next picture if he found no suitable definition.

D. Result experiment 2

The first variant of the qRANK was at this time not played as often as originally expected, so that no term was selected more often than twice. This value was too small to be a statement about the relevance of a tag. The second variant of qRANK was played more often and provided due to the small amount of data desirable results. An evaluation of the ranked tag list of the one hundred pictures provided, showed that 56 of the pictures had their most relevant tags in the first place. Only nine pictures did not have their most used tags in the first four positions (see **Figure 16**):

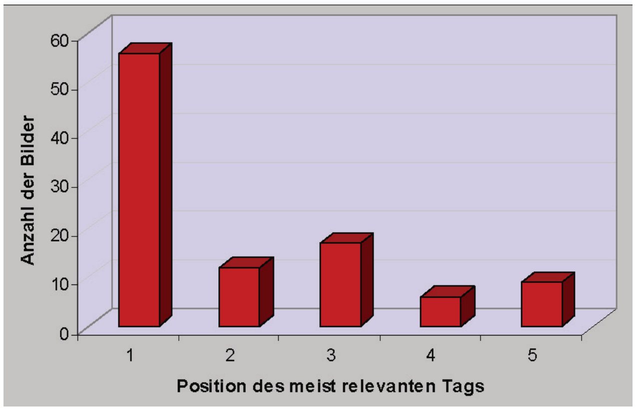

Figure 16. Positions of the most relevant tags

A one-week game period brought the result that **91% of the images that were played** during this time, had their most relevant tags **to the first four positions**. These results illustrate the effect of the approach. The aim of this experiment is not necessarily to find new terms for an image, but to assess the relevance of the existing tags depending on the content of the image. To make a useful statement only images were considered, that were selected at least four times. Striking here was that users often choose terms in different languages or plural/singular relation. So many images in the ranked tag list appeared often in different languages. Regarding the search process, this is not necessarily a disadvantage, since users search more multilingual. For the evaluation of the concepts in qRANK it is disadvantageous in the long run, as these terms are more preferred, and thus reduce the probability that other terms are selected. This problem can be limited, if we determine these relationships before automatically.

E. Resume

All over, information quality enhancement is getting more and more important – especially regarding the flood of autonomous Web resources without responding authorship. We presented exemplary the role of information quality in web-based information and knowledge transfer with smart interaction.

We adapted an existing assessment model to our purpose in qKAI and showed some examples for enhanced, rating-based interaction that is suitable to qualify Open Content stepwise in an incentive way. Incentive for user participation and interaction is implemented in qKAI as game-oriented, ontology-based and global rewarding model for any kind of interaction. Information quality can be utilized as a tool to derive personalization and user preferences in web-based information and knowledge systems, because it offers a.o. metrics to determine the fitness for use of autonomous, distributed resources.

The evaluation of our group-ranking and the game-based assessing approach for Flickr images showed promising results and the contents' quality increased obviously. Single tasks are reusable and combinable in different scenarios (implemented as atomic Web services).

XI. Further Use Cases and Examples

A. qMAP: A geo-coded visualization of Open Content

With qMAP [24] we implemented a map-based user interface to query, select and edit interlinked web resources. qMAP (see **Figure 17**) allows the user to filter DBpedia [39] entries and related multimedia content like Flickr images [40], YouTube [44] videos or Last.fm music [46]. Thematically and geographically personalized knowledge views are possible. Knowledge gaming content can be also placed on the qMAP.

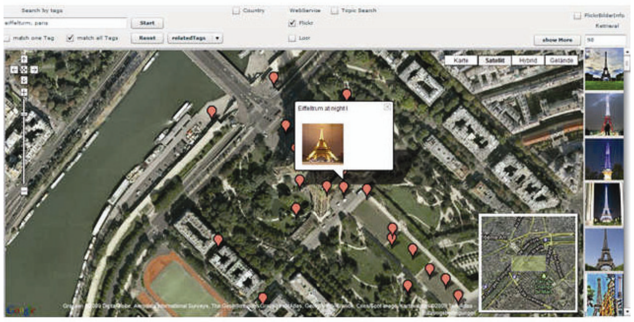

Figure 17. qMAP frontend

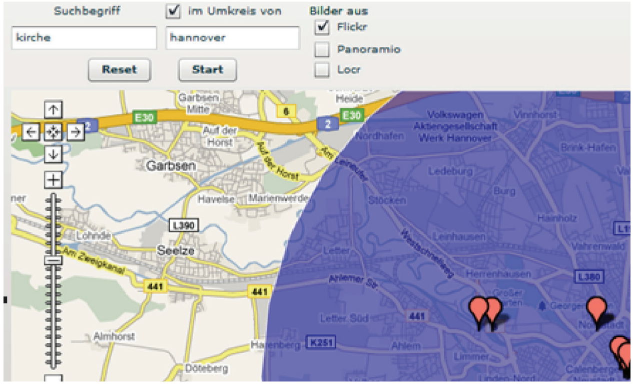

Figure 18. Search, filter and periphery interface of qMAP

Qualified Flickr images played first by qRANK are integrated into qMAP too (**Figure 17 and 18**). Figure 18 shows the periphery search and explore functionality of the qMAP. As shown in **Figure 19**, every user task and interaction is locked in qKAI's history protocol. Update, creation date or views of images are exemplary shown in Figure 19.

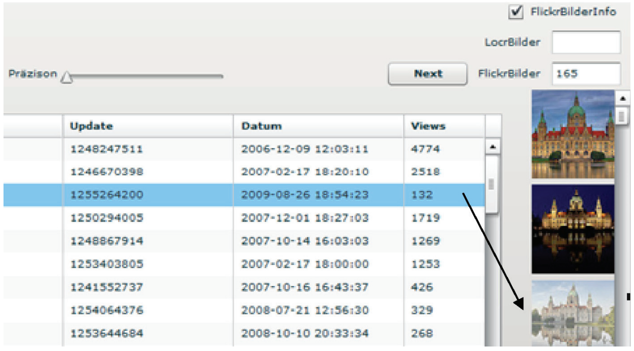

Figure 19. History and interaction protocol of Open Content for statistical analysis behind the qMAP interface.

The graphical interface of qMAP consists of three different states. So users can select with checkboxes individual functions or hide them. By default, a keyword-search in Flickr is set. The checkbox "*search by country*" the user can search for images within a certain radius and the checkbox "*topic search*" allows a search by topic. In order not to overload the map with markers, only a maximum of 100 images to each request is used. During the area search for Flickr images, the user has the additional option to set a radius (in km). The selected area is marked in blue on the map (see Figure 18).

B. *qMATCH: An assignment quiz with Flickr content*

qMATCH [48] is a prototype of an image-term assignment gaming type. First, the user enters a term he likes to get images about. Then he gets presented randomized terms and images out of Flickr and he has to assign the right term to the right image via Drag & Drop assignment (see **Figure 20**).

Figure 20. qMATCH text-image assignment game

Here we need a service called wrong-answerizer to assign wrong, but not stupid answers. Wrong-answerizer is deployed in further gaming types. qMATCH is useful to enhance e.g., language skills, geographically, architectural or historical knowledge. If we use term-term assignment a lot of vocabulary out of various domains can be assessed: assigning English to German translations, assigning buildings to right historical epochs or assigning cities to the right countries. In **Figure 21**, the statistically protocol of a user and his interaction on Open Content like Flickr images is shown.

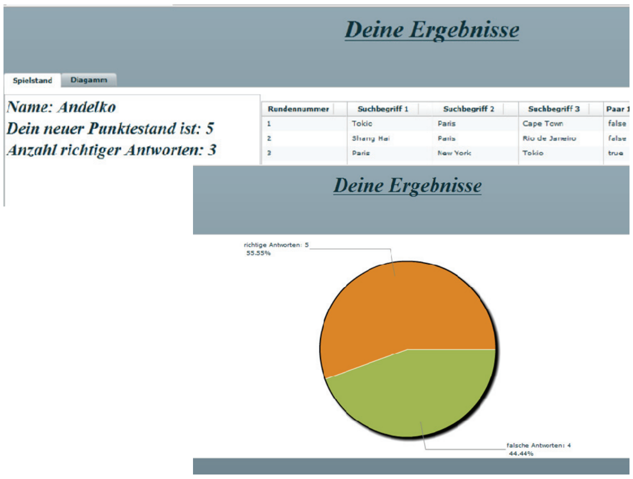

Figure 21. Knowledge game result in qMATCH with own correct answers and aggregated statistics.

XII. CONCLUSION AND OUTLOOK

We have exemplified the role of information quality in web-based information and knowledge transfer with smart interaction. Beyond evaluating the state of the art, we adapted an existing assessment model to our purpose in qKAI and showed some examples for enhanced rating-based interaction that is suitable to **qualify Open Content** stepwise in an incentive way. Incentive for user participation and interaction is implemented in qKAI as ontology-based, **global interaction rewarding system** for any kind of interaction (GIAR). All over, information quality enhancement is getting more and more important – especially regarding autonomous Web resources. Information quality can be utilized as a tool to derive personalization and user preferences in web-based information and knowledge systems, because it offers metrics to determine the fitness for use of autonomous, distributed resources.

The quality of the image search on the Web is a very topical subject of research. Many approaches and algorithms try to optimize the search. In this study, some possibilities are discussed and we implemented a **tag-based group ranking** method and a **game-based application** for the ranking of images. To show the effect of the procedure, images from Flickr were used. The focus of this contribution was the evaluation of user-generated metadata, which are derived from online communities, the so-called **tagging systems**. Especially for the search of images they are very important, because images, in contrast to other distributed content on the Web, do not contain metadata and are therefore difficult to find.

The simple form of tagging systems - free of any notation and relation of metadata generation - allows that content can be categorized by non experts. This, however, offers new challenges for Web search and data mining. The basic problem is to assess the relevance of the determined information. The advantage of the Semantic Web is that the information is in a machine-interpretable form because they were previously annotated semantically. It is different with metadata derived from the social annotation. Social annotation also called **collaborative tagging** arises when the common folk describe resources with keywords. In research, these are also known as **folksonomies**. To view the information from the folksonomies as useful advantage, they must be enriched with semantics. One possibility is to map them into lightweight ontologies. In this work, we discussed in detail how to combine folksonomies and tag ranking methods for images. The derived **keyword-oriented group search algorithms** and the **ranking game qRANK** are very promising, if the users are motivated to participate. Despite some weaknesses, tags are a useful addition to existing ontologies.

REFERENCES

[1] Steinberg, M. and Brehm, J.: Towards enhanced user interaction to qualify Web resources for higher-layered applications, Proc. DigitalWorld 2010, International Conference on Mobile, Hybrid, and On-line Learning (IARIA's eLmL), Neth. Antilles, ISBN: 978-0-7695-3955-3, pp.105-110, 2010.

[2] Naumann, F.: Quality-Driven Query Answering for Integrated Information Systems, Lecture Notes in Computer Science, Vol. 2261, Springer, 2002.

[3] Steinberg, M. and Brehm, J.: Towards utilizing Open Data for interactive knowledge transfer, Proc. DigitalWorld 2009, International Conference on Mobile, Hybrid, and On-line Learning (IARIA's eLmL), IEEE Press, 2009, pp.61-66, doi:10.1109/eLmL.2009.13.

[4] Kruse, P.; Warnke, T.; Dittler, A.; and Gebel, T.: Wertewelt Medien, http://www.nextpractice.de/fileadmin/studien/medienstudie2007/Medienstudie_Nov2007.pdf, 2007.

[5] Open Knowledge Foundation, The Open Knowledge Definition, http://opendefinition.org/, last update: 2008, visited: 2011-01-12.

[6] Steinberg, M. and Brehm, J.: Social educational games based on Open Content, Proc. International Conference on Intelligent Networking and Collaborative Systems (INCoS), Spain, 2009.

[7] Juran, J.: The Quality Control Handbook. McGraw-Hill, New York, 3rd edition, 1974.

[8] Resource Description Framework, http://www.w3.org/RDF/, last update: 2009, visited: 2011-01-12.

[9] Wand, Y. and Wang. R.: Anchoring Data Quality Dimensions in Ontological Foundations, Communications of the ACM, 39(11):86–95, 1996.

[10] Wikipedia, www.wikipedia.en, last update: 2009, visited: 2011-01-12.

[11] Freebase, www.freebase.com, last update: 2009, visited: 2011-01-12.

[12] Amazon, www.amazon.com, last update: 2009, visited: 2011-01-12.

[13] Bizer, C.: Quality-Driven Information Filtering in the Context of Web-Based Information Systems, Dissertation, 2007.

[14] Heath, T. and Motta, E.: Revyu.com: A Reviewing and Rating Site for the Web of Data, Proc. ISWC 2007, International Semantic Web Conference, Lecture Notes in Computer Science 4825 Springer 2007, pp. 895-902.

[15] Mui, L.: Computational Models of Trust and Reputation: Agents, Evolutionary Games and Social Networks, Dissertation, 2003.

[16] Parker, M.; Moleshe, V.; De La Harpe, R.; and Wills, G.: An evaluation of Information quality frameworks for the World Wide Web, Cape Peninsula University of Technology, University of Southampton, http://de.scientificcommons.org/14463068, 2006.

[17] IMS/QTI, http://www.imsglobal.org/question/, IMS Global Learning Consortium, Inc., last update: 2008, visited: 2011-01-12.

[18] Jøsang, A.; Ismail, R.; and Boyd, C.: A Survey of Trust and Reputation Systems for Online Service Provision, 2006.

[19] Nov, O.: What motivates Wikipedians? Communications ACM, Vol. 50, Nr. 11, 2007.

[20] Aperture, http://aperture.sourceforge.net/, Aduna, DFKI, last update: 2008, visited: 2011-01-12.

[21] ISO 15836: 2003, Information and Documentation – The Dublin Core Metadata Element Set, International Organization for Standardization, 2003.

[22] Openlink Virtuoso, http://virtuoso.openlinksw.com/, last update: 2008, visited: 2011-01-12.

[23] OpenNLP, http://opennlp.sourceforge.net/, last update: 2008, visited: 2011-01-12.

[24] Sarioglu, O.: Design and implementation of a map-based frontend with geocoded knowledge units, master thesis, Leibniz Universität Hannover, System- and Computer Architecture, 2009.

[25] Abbasi, R. and Staab, S.: Richvsm: Enriched vector space models for folksonomies, Proc. HT 09, Hypertext and hypermedia, pp. 219–228, New York, NY, USA, 2009. ACM.

[26] Marlow, C.; Naaman, M.; Boyd, D.; and Davis, M.: Ht06, tagging paper, taxonomy, flickr, academic article, to read, Proc. HT 06, Hypertext and hypermedia, pp. 31–40, New York, NY, USA, 2006. ACM.

[27] Panke, T.; Gaiser, S.; and Hampel, B.: Good Tags – Bad Tags, Waxmann, 2008.

[28] Wu, L.; Yang, L.; Yu, N.; and Hua, X.: Learning to tag. In 18th International World Wide Web Conference, pp. 361–371, April 2009.

[29] Stock, W. and Peters, I.: Folksonomies in Wissensrepräsentation und Information Retrieval. In Information – Wissenschaft und Praxis 59 (2008) 2, pp. 77–90, 2008.

[30] Vogel, A.; Anderson, A.; and Raghunathan, K.: Tagez: Flickr tag recommendation. 2008.

[31] Hepp, M.; Coenen, T.; and Van Damme, C.: Quality metrics for tags of broad folksonomies, pp. 118–125, Proc. International Conference on Semantic Systems, Journal of Universal Computer Science, 2008.

[32] Hotho, A.; Jäschke, R.; Schmitz, C.; and Stumme, G.: Folkrank: A ranking algorithm for folksonomies, Proc. FGIR 2006, 2006.

[33] Abel, F.; Henze, N.; and Krause, D.: Context-aware ranking algorithms in folksonomies, Proc. Webist, pp. 167–174, 2009.

[34] Peinado, V.; Artiles, J.; Gonzalo, J.; Barker, E.; and Ostenero, F. L.: FlickLing: a multilingual search interface for Flickr, Working Notes for the CLEF 2008 Workshop, 2008.

[35] San Pedro, J., Siersdorfer, S.: Ranking and classifying attractiveness of photos in folksonomies, Proc. WWW '09, 18th international conference on World Wide Web, pp. 771–780, New York, NY, USA, 2009. ACM.

[36] Li, X.; Snoek C.; and Worring, M.: Learning tag relevance by neighbor voting for social image retrieval, Proc. MIR '08, 1st ACM international conference on Multimedia information retrieval, pp. 180–187, New York, NY, USA, 2008. ACM.

[37] Garg, N. and Weber, I.: Personalized tag suggestion for flickr, Proc. WWW '08, 17th international conference on World Wide Web, pp. 1063–1064, New York, NY, USA, 2008. ACM.

[38] Negoescu, R. and Gatica-Perez, D.: Analyzing flickr groups, Proc. CIVR '08: International conference on Content-based image and video retrieval, pp. 417–426, New York, NY, USA, 2008. ACM.

[39] DBpedia, www.dbpedia.org, last update: 2009, visited: 2011-01-12.

[40] Flickr, www.flickr.com, last update: 2010, visited: 2011-01-12.

[41] Abel, F.; Henze, N.; Krause, D.; and Kriesell, M.: On the effect of group structures on ranking strategies in folksonomies, Weaving Services and People on the World Wide Web, pp. 275–300, 2008.

[42] Abel, F.; Frank, M.; Henze, N.; Krause, D.; Plappert, D.; Siehnde, P.: Groupme! - where semantic web meets web 2.0, pp. 871–878, 2008.

[43] Abel, F.; Henze, N.; Krause, D.: Groupme! In WWW, pp. 1147–1148, 2008.

[44] YouTube, www.youtube.com, last update: 2009, visited: 2011-01-12.

[45] Del.icio.us, http://delicious.com, last update: 2010, visited: 2011-01-12.

[46] Last.fm, www.last.fm, last update: 2009, visited: 2011-01-12.

[47] Wang, G. and Hoiem, D.: Learning Image Similarity from Flickr Groups Using Stochastic Intersection Kernel Machines, Proc. ICCV 2009.

[48] Jovancevic, A.: Analysis and extension of interaction with Open Content in the Social Semantic Web, master thesis, Leibniz Universität Hannover, System- and Computer Architecture, 2009.

[49] Cattuto, C.; Benz, D.; Hotho, A.; Stumme, G.: Semantic analysis of tag similarity measures in collaborative tagging systems; Proc. of the 3rd Workshop on Ontology Learning and Population (OLP3), July 2008.

[50] Specia, L. and Motta, E.: Integrating folksonomies with the semantic web, pp. 624–639, 2007.

[51] Sigurbjörnsson, B. and van Zwol, R.: Flickr tag recommendation based on collective knowledge, Proc. WWW '08, 17th international conference on World Wide Web, pages 327–336, New York, NY, USA, 2008, ACM.