



南京理工大学

Nanjing University of Science and Technology

Quality Evaluation of Social Tags according to Web Resource Types

Chengzhi Zhang & Li Lei

Content



南京理工大学

Nanjing University of Science and Technology

- Background
- Methods
- Results
- Conclusion and Future Work

Background



南京理工大学

Nanjing University of Science and Technology

- Users also annotate resources with different types of tags, such as objective tag, subjective tag, self-organized tag and so on.
- Because each web resource has its own characteristics, the tag types of each web resource are different.
- According to each web resource, the quality of each tag type is different.
- The usage of tag types and the quality of the tag can be different to different users and for various purposes of using tags.

Methods



南京理工大学

Nanjing University of Science and Technology

- Selected five types of annotated resources, namely blogs, books, images, music, and videos to start the research.

Methods

Table 1 Classification system of tag type with different annotated resources

Blog	vocabularies to describe the blog contents and showed up in the body
	vocabularies used in the blog title
	vocabularies used to describe blog type
	vocabularies used to self-organizing
	source of the blog (i.e. original, reprinted)
	self-feelings after reading the blog
	vocabularies to describe the blog contents, but did not show up in the body
	blog promulgator
	release time of the blog
	release location of the blog
Book	vocabularies to describe the book contents
	vocabularies to describe the book type
	vocabularies used to self-organizing
	vocabularies used in the book title
	self-feelings after reading the book
	book writer
	book language
	publish time
	publisher
	source of the book (knowing or getting the book, i.e. buying, getting as a present)
Image	vocabularies to describe the image contents
	vocabularies to self-organizing
	vocabularies to describe the image type
	vocabularies used in the image title
	self-feelings after watching the image
	source of the image (i.e. original, reprinted)
	release location of the image
	release time of the image
	image promulgator
	equipment used to take the image (i.e. Canon, Nikon)
Video	vocabularies to describe the video contents
	vocabularies used in the video title
	vocabularies to describe the video type
	vocabularies used to self-organizing
	self-feelings after watching the video
	leading role in the video
	director of the video
	source of the video (i.e. original, reprinted)
	video promulgator
	release location of the video
release time of the video	
Music	the singer
	vocabularies to describe the music type and genres
	vocabularies to describe the music contents
	vocabularies used in the music title
	vocabularies to self-organizing
	self-feelings after listening to the music
	language of the music
	release time of the music
ways of knowing the music (i.e. recommended by friends)	

Methods



南京理工大学

Nanjing University of Science and Technology

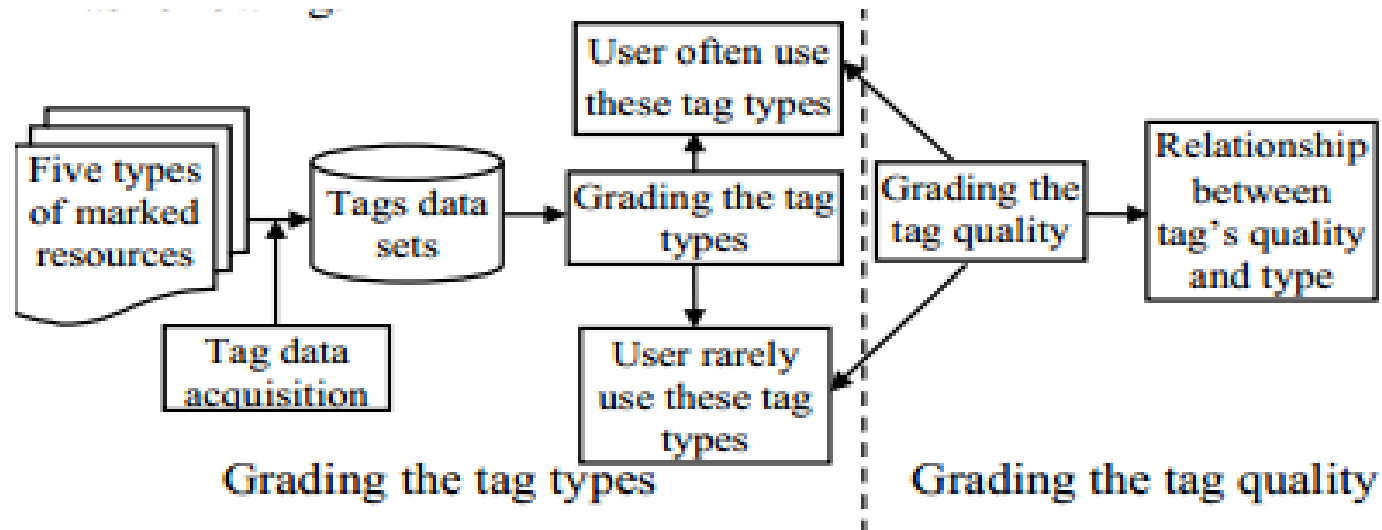


Figure 1. Discover the quality of each tag with different annotated web resources

Methods



南京理工大学

Nanjing University of Science and Technology

- Questionnaires Website is located at: www.sojump.com/jq/2709468.aspx.
- We have received 721 questionnaires in total, including 707 valid questionnaires. The effective rate is 98%.
- The questions were designed as whether users use these tag types to annotated resources or not.
- The user could choose Never, Occasionally, Sometimes, Often, and Always as the answer, given 1, 2, 3, 4, 5 points respectively according to his/her answer.

Table 3 The results of credit analysis and validity test

The dimension of the scale	Cronbach α	KMO	Bartlett sphericity test.
blog	0.790	0.787	2268.903
book	0.802	0.776	1167.392
image	0.802	0.720	453.697
video	0.865	0.804	1403.636
music	0.793	0.712	523.676

Methods



南京理工大学

Nanjing University of Science and Technology

- We researched on whether users' favorite tag types are the type of high quality or not.
- On the basis of five resources types we classified the type and graded the tag quality of two annotated web resources respectively which users are very familiar with.
- To ensure the tag quality score credibility, we found 31 users who have the experience of using tags to grade each two annotated web resources of these five resources types.
- Then we worked out the average scores of each tag and counted out the scoring average and variance of the quality of different tag types. The number of tags for blog, books, images, videos and music in our investigation is 20, 26, 28, 24 and 22 respectively.


Methods



南京理工大学

Nanjing University of Science and Technology

Forbidden City Museum (Beijing)



Museum 1 2 3 4 5

Forbidden City 1 2 3 4 5

Meridian Gate 1 2 3 4 5

Inner Golden Water Bridge 1 2 3 4 5

Square of Gate of Supreme Harmony 1 2 3 4 5

.....

Figure 2. Scoring sample for tags quality



Tag Type of Different Web Resources

- It can be seen that: except that users often use the singer's name as the musical marker mostly, users often use the words describing the resource content; the words appeared in the title or describe the resources type are used mostly by users; the words used for self-organization also are placed highly. Then the users can organize the resources according to their own necessity. The tag types that users seldom used are those tags describing the external information of the resources.

Results



南京理工大学

Nanjing University of Science and Technology

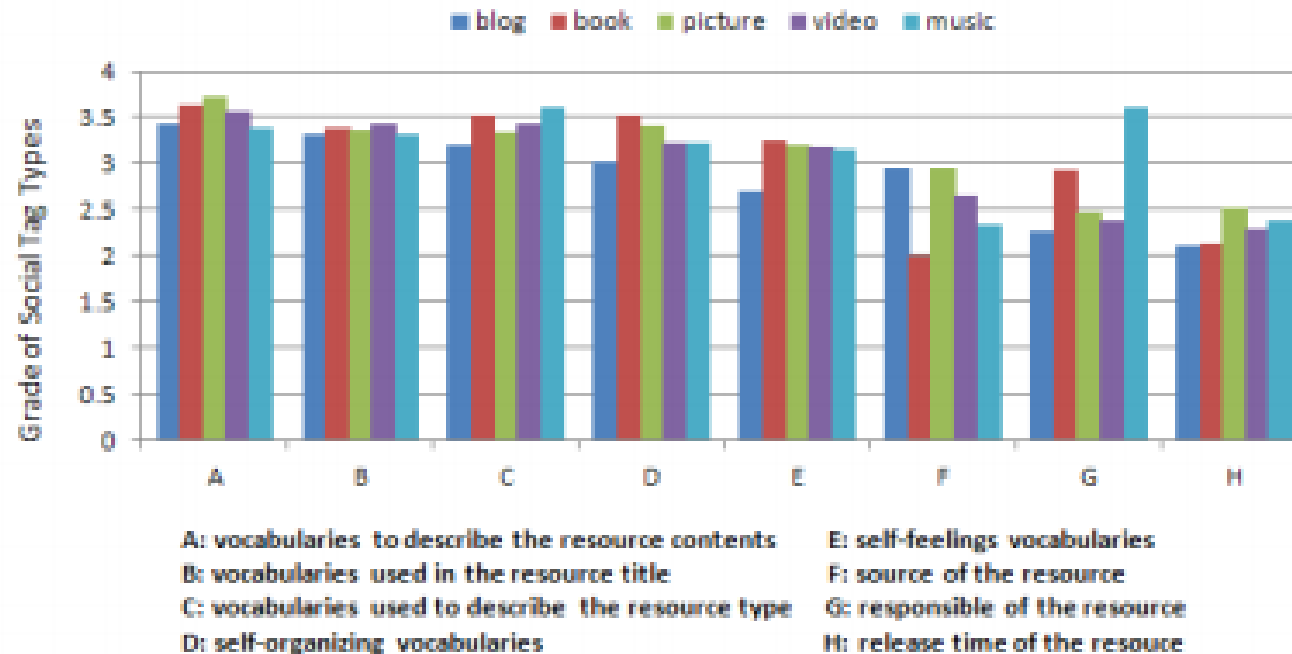


Figure 3. Comparison of Tag Types for Each Web Resource

Results



南京理工大学

Nanjing University of Science and Technology

The quality of different Web resources

- ✓ Above all, for each resource type, the words in the title, describing the resource content and type have the high tag quality and the tags that describe the external information have low quality scores which accord with the tag types used that users preferred in the questionnaire.
- ✓ But different resources have different characteristic and don't accord with the questionnaire. For example, for blog, the tags describing the blog content but not in the text have high scores; for book, the tags containing the author of the book have high scores; for image, the tags containing the publishing place of the image have high quality; for video, the tags containing the leading roles and directors have high quality. At the same time, we can see that even in one type of resources, the quality of each tag type of one resource type is different.



Conclusion and Future Work

- ✓ we firstly selected five web resource types, namely the blog, the book, the image, the music and the video, to explore the tag type used by users when annotating different resources.
- ✓ Then we selected specific resource and tags to explore the quality of each tag type according to these five resources and study the relationship between tag type and quality.
- ✓ In conclusion, we can see that the quality of tag types for different web resources is different. When evaluating tags quality, we need to combine specific web resource types according to the application scenes.
- ✓ In the future we will carry out the social tag quality evaluation of each web resource to generate the tag quality evaluation models according to web resource types and the users' need.

THANKS