



**HAL**  
open science

## Adding semantics to comics using a crowdsourcing approach

Mihnea Tufis, Jean-Gabriel Ganascia

► **To cite this version:**

Mihnea Tufis, Jean-Gabriel Ganascia. Adding semantics to comics using a crowdsourcing approach. 2016. hal-01288528

**HAL Id: hal-01288528**

**<https://hal.science/hal-01288528>**

Preprint submitted on 16 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Adding semantics to comics using a crowdsourcing approach

Mihnea Tufis

Universite Pierre et Marie Curie  
4 Place Jussieu  
75005 Paris, France  
mihnea.tufis@lip6.fr

Jean-Gabriel Ganascia

Universite Pierre et Marie Curie  
4 Place Jussieu  
75005 Paris, France  
jean-gabriel.ganascia@lip6.fr

## ABSTRACT

With over 85 million print units sold last year only for the top 300 comic book titles, the comics industry is reaching a new high for the first time since 2007 (before the economical crisis). And this excludes the increasingly popular graphic novels or the increasingly more accessible digital comics. Digital comics however find themselves at a point where the difficulty of producing high quality content is directly related to the lack of semantic information linked to it. In this short paper, we propose a crowdsourcing approach for annotating comic books for the purpose of producing XML-based encodings to assist publishers, researchers or collection curators.

## 1. INTRODUCTION

Most current comic book formats don't go beyond the simple packaging of resources and metadata in comic book archive files (essentially a regular archive file; e.g., CBZ is a ZIP archive, CBR is a RAR archive) that different reading systems un-package and commonly render on a page-by-page basis. Others (e.g., comiXology, Aquafadas) have taken a step forward in meeting the increasing demands of consumption of comic books on mobile screen devices (smartphones, tablets, phablets) and have introduced a panel-to-panel type of navigation. Critics of the latter are complaining though about the inaccuracy of the rendering, the loss of the contextual information that a classic comic book reader has by examining the proximity of the current panel and even about the fact that the original artistic intention of the creator is basically violated.

In this paper we discuss the use of crowdsourcing for gathering annotations for comic books content. We believe that the fan-base comics have and the increased consumption of digital comics will enable us to tap into a large, passionate crowd with deep and diverse knowledge about the genre. As a result, we are generating a corpus of structured metadata as Comic Books Markup Language (ComicsML) encodings for each annotated page. We discuss their potential use in the publishing industry (for digital comics), in digital humanities research as well as in curating comics collections and databases.

## 2. COMICS AND DIGITAL FORMATS

A synthetic comparison of e-book formats<sup>1</sup> reveals the multitude of digital publication formats available today and enumerates their capabilities with respect to what can be considered the key factors in terms of rendering digital comics: support for image, sound, interactivity, embedded annotations, word-wrap support or digital rights management. The linked table shows us that the most

compelling case seems to be made by the EPUB format. In addition to this, EPUB seems to be the digital publication format supported by most mobile platforms (with Amazon's Kindle being the only one not integrating it).

The ever-growing popularity of EPUB as a publishing open standard for e-books has brought a series of developments concerning image based publications (i.e. magazines, atlases, art albums etc.) with a focus on comics. The latest come from the EPUB Advanced and Hybrid Layouts working group (AHL) in the form of one of their newest specifications – Region Based Navigation [4]. This is essentially generalizing the guided navigation by describing a way to break the page of a comic book into different regions of interest (ROI) as well as their rendering order. In addition to this, a Metadata Structural Vocabulary for Comics [8] is describing a minimal set of terms for annotating regions of interest: panel, panel-group, balloon, text-area and sound-area. The primary goal is assisting creators of digital content in using the combined metadata and reading system capabilities for an enhanced rendering of the content.

While tremendously helpful, the Region Based Navigation specification is taking care exclusively of the presentation layer (i.e. rendering a publication on a screen device). We believe that the artistic nature of classic comics as well as the great potential digital comics have already showcased allow us to go beyond simple content presentation. Comics have always engaged their readers and have always drawn passionate people not only into reading them. Comic books aficionados are proud of their collections, they establish an intimate link with their characters – they want to learn their history (e.g., when did Batman appear for the first time, in which issue, which cartoonist drew him?), they follow their lives (e.g., how many times did Wolverine die, did Superman and Spiderman ever cross paths, where are the Watchmen meeting?). For all these cases and numerous others there is a clear need for richer descriptive metadata, which go beyond the presentation scope for which EPUB's Structural Vocabulary for Comics has been created.

## 3. CROWDSOURCING ANNOTATIONS FOR COMICS

We rely on the digital comic book readers as participants to our crowdsourcing experiment. The tasks we propose to them are organized around a set of questions regarding a series of problems related to comics on which computer algorithms are not performing well enough: page structure, identification of narration elements (characters, places, events), stylistic elements (bubble shapes, onomatopoeia, movement lines).

We aggregate the answers taking into account the reliability of a user in a given context (task difficulty, user experience with the task, type of question) and the agreement between the annotators.

---

<sup>1</sup> [http://en.wikipedia.org/wiki/Comparison\\_of\\_e-book\\_formats#Comparison\\_tables](http://en.wikipedia.org/wiki/Comparison_of_e-book_formats#Comparison_tables)

A quality score is thus generated for each annotation with the best of them being selected as solutions.

We subsequently are able to generate the ComicsML [17] encodings. This XML derived format is particularly useful since it's based on the already successful Text Encoding Initiative (TEI), allowing for declarations for page structure and composition, panels, characters, text (in all the varieties hosted by the comic medium: different types of balloons, diegetic text, onomatopoeia), events and even panel-to-panel transitions

First, it could enable comics publishers to generate better content: the ComicsML content is aggregated to the EPUB layer, functioning as a semantic complementary layer to the (mainly) presentation layer offered by EPUB. In addition to that, the collected annotations can be directly used within the EPUB Rendition Navigation document [4], according to the recommendations from the EPUB Structural Vocabulary for Comics [8].

Digital Humanities scholars could use such encodings to optimize their research on comics: structured ComicsML encodings of comics or graphic novels could be used to explore the content in a multitude of ways. This applies not only to digital comics, but also to ComicsML encodings of physical comics.

Comics curators and collectors (professional or amateur) of physical or online collections could be provided with a structured content which could be more easily integrated within their collections or databases. This may assist them into enlarging public or private databases of characters or comics series and enable the creation of artefacts such as comic books dictionaries, indices, and dictionaries of onomatopoeia. A certain number of projects are already in place and could greatly benefit from the creation of comic books annotations. We mention here the Grand Comics Database (an online database of printed comics), Comic Book Database, Digital Comics Museum (a collection of scanned public domain comics from the Golden Age) or the Catalogue of the *Cite Internationale de la Bande Dessinee et de l'Image* from Angouleme.

### 3.1 The crowd and the platform

With several hundred different crowdsourcing platforms already up and running, we turned our attention to the Citizen Science Data Factory comprehensive technical report [2] and their recommendations for platform design.

Then, a project that became a great source of inspiration was the tagger, from the Public Catalogue Foundation (PCF). Set to annotate the entire collection of paintings from UK museums, PCF's tagger is a great application of crowdsourcing in the field of arts. With their concept in mind, we wanted to take a step in a different direction – producing annotated content for publishers, researchers or curators to use.

When designing a crowdsourcing project or an entire platform there are several factors that absolutely need to be taken into account [14]:

- What crowd are we targeting?
- Subsequently, how do we motivate it? What incentives will we use?
- How will tasks be designed?

Concerning motivation we use exactly the same incentives as PCF's tagger – offering the possibility to passionate comic readers to offer their (potentially undiscovered) expertise while doing

something they love. Our hypothesis is that comic book readers are so passionate about their comics that they would pay a great deal of attention to details; giving them the opportunity to showcase their “skills” should be a strong enough incentive. In addition to that, offering step-by-step instructions, examples and tying them to the comic book theory might actually get non-frequent readers more involved into the comic books world. For all we know, it's a virus that catches on fast!

Our platform needed to be simple, yet offer us the possibility of creating a high quality user experience. The state of the art research pointed us to Crowdcrafting – a web platform dedicated to creating, managing and running crowdsourcing projects. Developed over PyBossa, a BOSSA API for crowdsourcing engines, the platform has a series of advantages: i) easy authentication (also available through 3rd party sites: Google, Facebook) – great feature, meant to get everyone started fast and prevent potential users from getting bored before even diving into the tasks [3], ii) the possibility to configure the user interface using open web technologies - the UI has been entirely developed in HTML and JavaScript (the use of Bootstrap is encouraged!), iii) the tasks are described using JSON objects or CSV files, and iv) the answers are stored as JSON objects. Besides the actual annotation, PyBossa enables us to collect useful information about the users working on each task and the time they are spending on it. This in turn allows us to develop our quality evaluation algorithm, which aggregates the annotations based on contextual information about both the task and the users.

### 3.2 General task design

Before designing our tasks we needed to elicit a set of requirements, based on the needs of the potential stakeholders (readers, publishers, researchers). Thus, the following aspects were taken into account. First, we looked at the limits of digital formats for comics (notably EPUB), which we discussed in Section 1. Second, we considered the ComicsML schema as a strong reference for some of the metadata we'd like to collect. Finally, principles of the comic book theory emerging from the work of masters of the genre such as Will Eisner's instructional books [6] and Scott McCloud's scholarly work in “*Understanding Comics*” [11] provided us with a greater understanding of the sequential art in general. We have designed our tasks with attention to concepts such as page composition, page structure, reading behaviour, ellipsis and transitions or the interplay between text and visual.

Structurally each task asks the annotator to answer a series of sequential questions about a comic book page. The questions are centred on 4 key themes (Figure 2).

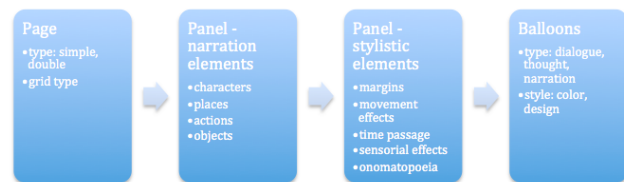


Figure 1. The 4 key annotation themes

Considering the theory related to the unity of the panels of a comic book page [11] and to prevent annotators to spend too much time on a task [3], we decided to centre our tasks on the entire page, as opposed to presenting them with each panel. Also in order to keep the process fast and simple as in the case of the PCF tagger, the annotators are not allowed to return to a previous step.

### 3.3 Task description

#### 3.3.1 Step 1 – Page structure

The annotators are presented with a simple interface in which they will have to make a choice between a set of suggested grid layouts. These layouts are the output of applying the automatic frame extraction algorithm developed by Rigaud et al. [14].

In addition, our mobile annotation app (see Figure 2) also allows the annotators to draw the grid layout as they wish, by tapping over the presented page.

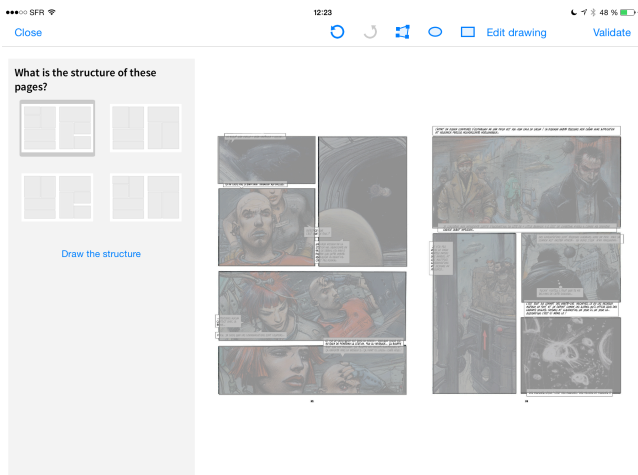


Figure 2. UI for Step 1 - Page Structure

#### 3.3.2 Step 2 – Panel reading order

Sometimes the reading order of the panels of a page is clear. This happens in older comics in which a regular (e.g., the classic 3x3) grid is fairly obvious to read. As the comics evolved as a genre, so did the artistic means become more and more sophisticated. It is not rare to encounter difficult to read grids, as a way for the artist to challenge and engage the reader. At this stage we apply the previously identified grid over the presented page and we ask the annotator to tap / click over each panel in the reading order.

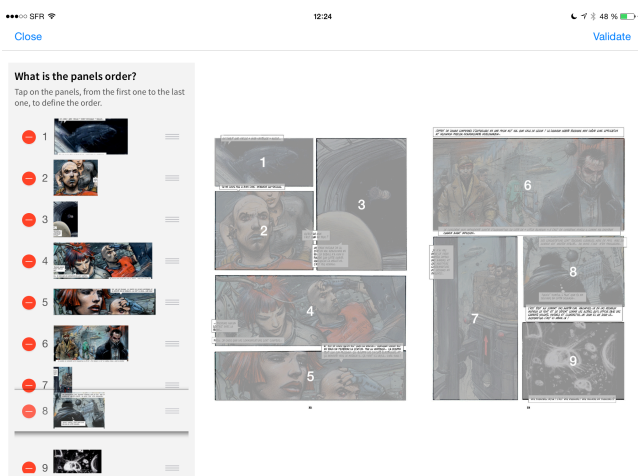


Figure 3. Panels reading order

This step is particularly important since we use the difficulty of interpreting a page as a proxy for the difficulty of the task. We compare it with the time the contributor is spending on the task and with the aggregate answers of the other annotators on the same task; thus we establish the quality of the current annotator in

the given context and the quality of his answer.

Finally, considering the answers of steps 1 and 2 we can already initialise the ComicsML encoding of the current page with the tags related to the page structure and composition (i.e. insert the `cbml:panel` elements in the given order).

#### 3.3.3 Step 3 – Character identification

At this step, we ask the crowd to simply enumerate all the characters they can identify in the current page. Characters can be identified by reading their names in the text, recognising them from experience or simply giving a general statement about the character (e.g., “*masked man*” may be referring to *Batman*). The difficulty here is that the page is taken out of the context of the whole comic book that might make it difficult for an annotator to name a secondary or referential character whose name has been mentioned on a different page. At this point we will be using state of the art symbolic learning algorithms to fusion generic with specific information (e.g. if *Batman* and “*masked man*” both have high quality scores, they will both denote the same concept and will be considered as valid annotations). In this case, upon detection of a potential fusion, our algorithm will suggest the more specific term, leaving the final word to the human annotator.

#### 3.3.4 Step 4 – Places identification

At this step the annotators are asked to simply enumerate all the places they can recognise on the current page. We are particularly looking for named places (e.g., *Gotham City, NY, planet Mars*), but will also ask the annotator to signal any generic place that he might deem as important for the scenes in the page (e.g., “*the interior of a bank*” [in case of a robbery], “*inside a space ship*” [in case of a space battle]). These are exactly the kind of very specific annotation tasks for which state of the art image recognition algorithms are expected to fail.

#### 3.3.5 Step 5 – Events identification

This is yet another highly specific recognition task. Annotators are asked to describe the most important actions occurring in the page, or if they are willing the actions taking place in each of the panels. At this stage we expect a heterogeneous corpus of annotations and we have yet to figure out what techniques will we use to validate and finally aggregate them. This is a pivotal step, since the solutions generated here, together with the annotations obtained in steps 3 and 4 will be used to further build the ComicsML encoding of the page (i.e. inserting the `characters` attribute of the `cbml:panel` element, the `events` element, the `places` element).

Scott McCloud stresses the role of ellipsis (“the blood in the gutters” – the space between two panels, where the reader fills in the gaps between the events depicted in two adjacent panels) as an artistic mean for authors to engage their readers and describes a typology of these transitions [11]. It turns out ComicsML allows us to declare such transitions through the `#ana` attribute of the `cbml:panel` element. Our ambition is to use the events descriptions for each panel and try to automatically infer the type of transition between successive panels. This might prove to be of utmost importance to digital comics publishers (they exploit the ellipsis more and more, by adding different transition effects, animations and transformations between panels and pages) as well as comics scholars working in digital humanities (e.g. it is well known that traditional comics from different geographic areas - Japanese manga, American comics, Franco-Belgian bande dessinée - have different frequency distributions for each type of such transitions).

### 3.3.6 Step 6 – Non-visual cues

Comics are a special medium, making use of the visual to depict all other non-visual senses, by making use of different drawing tricks:

- Smoke coming out of a cigarette may engage the reader’s smelling sense
- Onomatopoeia form a particular language of their own; comics and especially manga authors have proven to be extremely creative when it comes to expressing different sounds via stylised text (e.g., “*POW!*” can be a punch, “*BAM!*” can be a gun being fired, the Japanese manga has developed a complex vocabulary of its own)
- Horizontal lines around a car can suggest the car is moving at high speed, while around a ball they express the simple movement of the ball

At this step, we ask the users to identify precisely these 3 types of non-visual sensations and the way they are expressed: smell, sound and movement.

Digital comics publishers could largely benefit from this kind of annotations, by adding visual and sound effects to their content [3]. On the other hand, researchers could study, for instance, the drawing style of an author and his use of non-visual cues, and go as far as creating onomatopoeia dictionaries for comics (to our knowledge, such dictionaries already exist for manga, but not for American or European comics; curiously enough the complexity of the meaning of the onomatopoeia in Japanese manga far exceeds that of its American and European counterparts).

### 3.3.7 Step 7 – Putting everything together

At this step, annotators are presented with all the previously identified narrative elements (characters, places, non-visual cues) and are asked to simply drag and drop them over the panels in which they appear.

At the end of this stage, we should be able to generate a reasonably complete ComicsML encoding of the current page (see Figure 3).

```
<cbml:panel
  n="2"
  characters="#pilot #leyla"
  ana="#subject-to-subject"
  xml:id="case_002"
  xmlns:cbml="http://www.cbml.org/ns/1.0">
  <cbml:balloon xml:id="boule_002" type="speech" who="#pilot">
    You should reconsider. Look above!
  </cbml:balloon>
  <cbml:balloon type="speech" who="#leyla">
    What is this thing?
  </cbml:balloon>
</cbml:panel>
```

Figure 4. A fragment of the ComicsML encoding for the page presented above

## 3.4 A note on the quality of annotations

In this section we briefly discuss our method for assessing the quality of crowdsourced annotations, prior to their aggregation.

To this respect, recent body of work in the field of crowdsourcing focuses on methodologies, factors and measures for crowdsourced annotations. The problem seems to be that there isn’t a one size fits all approach. Not for developing, nor for assessing the success of a project. Each project, depending on the crowdsourced tasks requires tinkering with at least two parameters: the size of the crowd and the time required for the project to converge to satisfactory results.

Alonso and Mizzaro [1] use K-statistics to measure inter-annotator agreement (IAA) and compare experts to non-experts on binary relevance tasks, Snow et al. [16] obtain good results with the same kind of measures for 4 different types of tasks (affect recognition, word similarity, textual entailment, word sense disambiguation), but do encourage to model the bias of individual annotators and correct it. Feng et al. [7] as well as Kazai and Milic-Frayling [10] work on relevance assessment tasks and use inter-annotator agreement measures, but while the former suggests the removal of outlier annotators, the latter propose an approach for modelling trust for annotators as a function of the IAA, their practical involvement in offering comments to tasks and the degree to which they accumulate knowledge while performing tasks.

Sheng et al. [15] use the majority voting system and try to model the uncertainty associated to each label to determine whether further labelling is required or not; however, in response to their strict assumptions about their crowd, Donmez et al. [5] suggest to work with unknown label accuracies, and in turn estimate the accuracy of the labellers.

Considering crowdsourcing image annotations, Nowak and Lukashevich [13] use predictors in order to place a new annotation in a ontology of concepts and then measure the ontology distance between the annotation and a gold standard corresponding concept. Later, Nowak and Ruger [12] combine different evaluation metrics (ontology score, equal error rate, area under the curve, ranking correlation) with classic inter-annotator agreement and conclude that while generally reliable the IAA results are not consistent over the different evaluation metrics. According to them, finding a more robust quality measure for measuring agreement in image annotation tasks is still an open research question.

We base our approach on the idea that each annotator performs differently on different tasks according to his motivation, level of expertise and task difficulty. Moreover, we expect annotators to become better in time – we expect an increase in the level of expertise for the different types of tasks.

The quality of an annotator  $a_i$  on task  $t$  is a function of the difficulty of task  $t$  and expertise of annotator  $a_i$  on the task.

$$Q_{annotator}(a_i, t) = f(\text{difficulty}(t), \text{expertise}(a_i, t))$$

The quality of the annotations corresponding to task  $t$  is given by the inter-annotator agreement between annotators from  $A$ , working on task  $t$ .

$$Q_{annotation}(t) = IAA(A, t),$$

where  $A = \{a_j, Q_{annotator}(a_j, t) > s\}$  is the set of annotators for which the annotator quality on task  $t$  is above a certain threshold  $s$ .

## 4. CONCLUSIONS AND FUTURE WORK

The practical motivations behind our work can first be found in the increased market share of comics in the publishing industry (and the constant increase of digital comics there within). Second, taking into account the increased interest from the part of digital humanities scholars with respect to comics and graphic novels as well as the special interest collection curators (digital or physical, amateurs or professionals) take into collecting, classifying or searching comic books corpora, we have stressed the lack of, and therefore necessity for a standardised, digital encoding format for comic books, graphic novels, manga or bandes dessinées.



It is difficult however to develop automatic information retrieval solutions to populate such ComicsML encodings. We have thus shown how crowdsourcing the resulting tasks over the web, we can involve a potentially large number of a-priori non-expert annotators, working on their passion and expertise with reading and experiencing comics.

We have presented the general architecture of our system, as well as details of how we have designed our tasks, having in mind three main aspects: the limits of current digital comic book formats, the specifications behind the ComicsML metadata schema and theoretical principles of the comics as a genre. For each step we have discussed the challenges we have faced or will be facing as well as the way the collected results will merge into the final ComicsML encoding.

Finally, we have brought to our readers' attention the sensitive topic of crowdsourced data quality. Thus, based on state of the art research we have introduced the general idea behind our approach to annotation quality: estimating context-dependent annotator quality, which in turn is used to determine the quality score for a given annotation.

#### 4.1 Future Work

At the time this paper was submitted, we have finished the prototype development for both the mobile app and the web application. At the same time, we are closing in on refining our data quality algorithm that should be ready in the following month.

Considering the corpus of comic book pages, we are currently negotiating the use of copyright-free comics from the online Digital Comics Museum, as well as licensed digital comics from one of our commercial partners.

Based on the approach of Feng et al. [7] our methodology for collecting data will comprise of two steps: a calibration step with the purpose of establishing tasks typologies, the number of annotators required for each tasks as well as enough testing data to calibrate our quality algorithm should be released before June 2015; the second, large-scale release is to be launched before September 2015. However, the first ComicsML prototype encodings and hopefully the first hybrid EPUB+ComicsML digital comic book should be ready as early as the summer of 2015.

## 5. REFERENCES

- [1] Alonso, O., Mizzaro, S. Using Crowdsourcing for TREC assessment. *Information Processing & Management* 48, pp. 1053-1066, 6 (2012).
- [2] Azavea, SciStarter. Citizen Science Data Factory – Part 1: Data Collection Platform Evaluation. 2014.
- [3] Cognito Comics. *CIA: Operation Ajax*, 2012. <http://www.cognitocomics.com/operationajax/>.
- [4] Conboy G., Duga, B., Gardeur, H., Kanai, T., Kopp, M., Kroupa, B., Lester, J., Garrish, M., Murata, M., O'Connor, E.

- EPUB Region Based Navigation 1.0*, 2014. <http://www.idpf.org/epub/renditions/region-nav/>.
- [5] Donmez, P., Carbonell G. J., Schneider, J. Efficiently Learning the Accuracy of Labeling Sources for Selective Sampling. In *Proc. 15<sup>th</sup> ACM SIGKDD*, ACM Press (2009), 259-268.
- [6] Eisner, W. *Comics and Sequential Art: Principles and Practices From the Legendary Cartoonist*. W.W. Norton & Company, USA & U.K., 1985.
- [7] Feng, D., Sveva, B., Zajac, R. Acquiring High Quality Non-Expert Knowledge from On-demand Workforce. In *Proc. Workshop on the People's Web Meets NLP 2009*, Association for Computational Linguistics (2009), 51-56.
- [8] Ichikawa, D., Kasdorf, B., Kopp, M., Kroupa, B. *EPUB Region Based Navigation Markup Guide 1.0*, 2014. <http://www.idpf.org/epub/guides/region-nav-markup/>.
- [9] Kazai, G., Milic-Frayling, N. On the Evaluation of the Quality of Relevance Assessments Collected through Crowdsourcing. In *Proc. SIGIR 2009 Workshop on Future of IR Evaluation*, ACM Press (2009).
- [10] McCloud, S. *Understanding Comics – The Invisible Art*. William Morrow – An Imprint of Harper Collins Publishers Inc., USA, 1993.
- [11] Nowak, S., Ruger, S. How reliable are annotations via crowdsourcing? A study about inter-annotator agreement for multi-label image annotation. In *Proc. MIR 2010*, ACM Press (2010), 557-566.
- [12] Nowak, S., Lukashevich, H. Multilabel Classification Evaluation using Ontology Information. In *Proc. IRMLes Workshop 2009*.
- [13] Rigaud, C., Tsopze, N., Burie, J.-C., Ogier, J.-M. Robust text and frame extraction from comic books. In *Proc. GREC 2011*, Springer (2011), 129-138.
- [14] Sharma, A. Crowdsourcing Critical Success Factor Model: Strategies to harness the collective intelligence of the crowd. *Working paper* (2010).
- [15] Sheng, V., Provost, F., Ipeirotis, P. Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers. In *Proc. 14<sup>th</sup> ACM SIGKDD 2008*, ACM Press (2008), 614-622.
- [16] Snow, R., O'Connor, B., Jurafsky, D., Ng, A.Y. Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. In *Proc. EMNLP 2008*, ACM Press (2008), 254-263.
- [17] Walsh, J.A. Comic Book Markup Language: an Introduction and Rationale. *Digital Humanities Quarterly* 6, 1 (2012). <http://www.digitalhumanities.org/dhq/vol/6/1/000117/000117.html>.

**Columns on Last Page Should Be Made As Close As Possible to Equal Length**