



**HAL**  
open science

# Ontologies and spatial relations applied to comic books reading.

Clément Guérin

► **To cite this version:**

Clément Guérin. Ontologies and spatial relations applied to comic books reading.. PhD Symposium of Knowledge Engineering and Knowledge Management (EKAW), Oct 2012, Galway, Ireland. hal-00743480

**HAL Id: hal-00743480**

**<https://hal.science/hal-00743480>**

Submitted on 19 Oct 2012

**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.

# Ontologies and spatial relations applied to comic books reading

Clément Guérin

L3I, University of La Rochelle, France  
clement.guerin@univ-lr.fr

**Abstract.** In this paper, we present a procedure to add semantics to digitized comic books using ontologies. We focus on the panels, computing the reading order thanks to their spatial relations.

**Keywords:** Ontology, Image, Comic Books, Spatial Relations.

## 1 Introduction

Since the last century, comic books have become an important piece of the cultural heritage for a lot of countries. Western europe, but also USA and Japan respectively with comics and mangas, have produced a large amount of material which is now massively digitized according to preservation and content enhancement concerns. However, the digitisation is often limited to the page as a whole without digging any further in the very content of that page. If we were able to go beyond the simple page digitisation we could imagine a whole new field of exploitation of the 9th art, such as smartphone-specific reading and data mining through a large set of albums. Dealing with comic books on digital hardware, *digital comics*, which are meant for a digital exploitation, must be distinguished from *digitized comics* that remain semanticless static pictures. Some industrial actors have already started to develop tools providing means to translate digitized comics into digital comics. This remains however a quite time consuming task due to the fact that everything, from the panel splitting to the sequence ordering, has to be done manually. The addition of a few pieces of semantic, such as panels ordering or associations between characters and speech balloons, could be a great deal to enhance the whole process of creation. Our work focuses actually on the bridge between digitized and digital comics by attempting to add this kind of semantics to physical documents and consequently, narrowing the semantic gap which is a major issue for the image analysis community.

A few research teams have conducted works on comics image processing [1,2,3] Unfortunately, these are often reduced to panels and balloons extraction, without concern about the associated semantic. Yet, comics contain scenaristic information, which are naturally embedded in the drawings, but also in the physical structure and topological ordering of the panels within the page. This knowledge can be represented by an ontological model providing a new and original support for navigation.

Indeed, ontologies [4] applied to the semantic web [5] allow to organize domain specific knowledge in order to share and store it. They can be used to describe

and define a structure for a domain specific knowledge representation. Regarding the image processing field, they allow to overcome the implicate signification of a visual information issue and can be used as a mutual reference to share semantic content.

This paper presents the elaboration of such an ontology and how it can be used for comic content retrieval and visualisation. The second part introduces the comic book's vocabulary while the third section describes our model applied to panels ordering. The last part concludes this work and bring on some prospects.

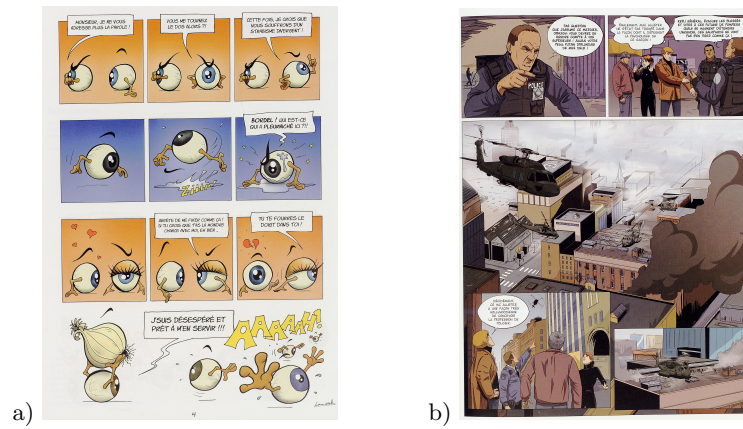
## 2 Comic books: pattern and vocabulary

The retrieval of semantics from images is a challenging task, especially with random pictures picked from unstructured databases. Working with comicbooks pictures presents the advantage to rely on a structure and a language that has already been studied [6] and that can be modeled. Comic books stories can be told during decades through a succession of volumes. Each of these volumes is made of pages containing panels which can be, but not necessarily, gathered into strips. The reading order is defined by pages order, which can be from left to right or right to left, depending on the origin of the material. Let *Comic* be an album where  $Comic = \{p_1, p_2, \dots, p_n\}$ , with  $n$  the total number of pages and  $p_i$  the  $i^{th}$  page of *Comic*. [7] defines the reading order as  $p_1 < p_2 < \dots < p_n$ .

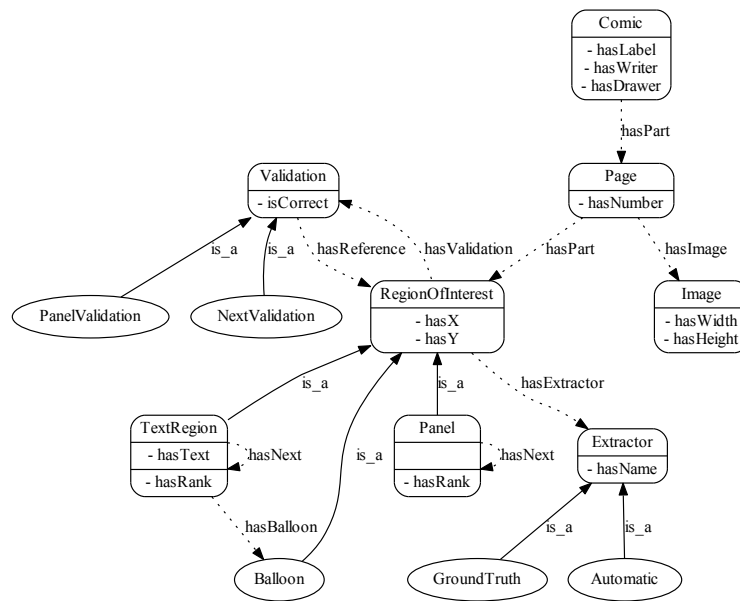
Each panel contains specific figures that, all together, define a scene. These figures can be characters but also text balloons, background, onomatopoeias and more generally any kind of drawn object. A panel has usually a rectangular shape, comes with or without border, and can be separated or not from other panels by gutters (see Fig. 1a). Sometimes, the panel's geometry can be more complex though. Indeed, there are a noticable amount of comic books, mostly from America, Japan and, for the most recent of them, Europe, where panels can contain one another or overlap each other. Let a page be  $p_i = \{f_1, f_2, \dots, f_m\}$  where  $m$  is the number of panels in this page and  $f_k$  its  $k^{th}$  panel. The reading order  $f_1 < f_2 < \dots < f_m$  relies both on spatial relations between these panels and the scenaristic interpretation of the reader. Spatial relations are not necessarily sufficient to determine whether a panel should be read before or after another. In fact, authors can introduce ambiguous situations on purpose to serve the screenplay. For instance in Fig. 1b, the insertion of a panel inside a bigger one creates a simultaneity effect in the scene [8]. Then, the page can be read following different paths and each reader will make his own choices.

Moreover, the reading order is different regarding the origin of the comic (for instance, japanese mangas are read from right to left). Generally speaking, authors are free to redefine the reading sequence as they wish and, even if the consensual reading order is from left to right and top down, litterature abounds of counterexamples. That is the reason why we choose to model the panels arrangement with a partial order, from which the reading order can be deduced using a topological sorting.

Most recent works [1,2] have introduced comic books processing algorithms. They aim to split up a digitized page in order to retrieve visual information



**Fig. 1.** a) Straightforward reading sequence, top-down and left-to-right. [9] b) The three panels at the bottom of this frame show the same scene from different angles. It is not obvious which one is meant to be read first. [10]



**Fig. 2.** Piece of our model concepts hierarchy (arrows) and their properties (dotted arrows).

such as panels, balloons and text. Strong methods, like *Hough transform*, *X-Y recursive algorithm* or gradient-based, are used to acheive this purpose but their efficiency often relies on a few parameters (panels border color, gutters color, balloons color, etc.). These last values are low-level pixel properties. Our work is focused on these works exploitation and interpretation.

### 3 Ontologies and comic books

#### 3.1 State of the art on image and comics applied ontologies

Many studies have been conducted on the issue of narrowing the semantic gap. For instance, [11] worked on the use of automatic learning methods (SVM, decision-trees, neural networks, bayesian network) in order to provide a way to classify images (or regions of interest) according to a set of pre-learned concepts. [12] merged visual, conceptual and contextual information in order to build the semantic hierarchy of an image. [13] used a description logic to understand scenes while [14] introduced humans in the annotation loop.

[7] and [15] focused on narrowing the semantic gap in the special area of comic books. [7] have build an ontology whose purpose is to identify regions of low interest in the images, in order to get rid of them when displayed on the small screen of a mobile phone. The ontology brought by [15] conceptualize the comic content extraction process but have no interest in its scenaristic facet.

#### 3.2 Global model

Our ontology is meant to achieve three purposes:

- Describe comic books at the image level. Knowledge will consist in regions of interest visual characteristics and characteristics of the image itself. This level will be labeled “low level”.
- Describe the scenaristic part of a digitized comic book. The focus will be put on the comic’s semantic. Every piece of knowledge deducable from the “low-level” will be described in this “high-level”.
- Extraction algorithms evaluation.

In order to be consistent with the content extraction model philosophy, our ontology is based upon the concept of *Region Of Interest* or *ROI*. A *ROI* is an image area considered as being somehow meaningful. It can be specialized into a *Panel* or a *TextRegion*. Each *ROI* is attached to the page it comes from, that page being associated with its album through the transitive relation *hasPart*. Each page is also linked to a digital image file by the *hasImage* relation.

A *ROI* is computed by and linked to a specific extractor following the *hasExtractor* relation. This extractor can be automatic, in that case it is an algorithm, or human when ROIs are extracted by a human being. The *Validation* concept and its specializations, connected to *ROI* by *hasValidation* and *hasReference* properties, allow us to compare different extractor’s results to each other.

Fig. 2 illustrates a part of our model<sup>1</sup> describing the present work.

<sup>1</sup> The complete model is available here:

<http://l3i.univ-larochelle.fr/IMG/zip/comicbooks.zip>

### 3.3 Populating rules

The ontology is populated with low level data provided by the different extractors. These data are used as a basis for the high level populating rules. The first piece of information to be deduced from these data is the panel ordering, which is represented in our model by the *hasNext* relation.

Each panel is characterized by a polygon which takes its coordinates into the page system reference, the origin being the top-left corner. Two panels,  $A$  and  $B$ , from a page are bounded by directionnal and topological spatial relations. Regarding the first ones, we have chosen to use three relations defined by RCC8 [16], namely disjonction  $DC$ , overlapping  $PO$  and inclusion, with and without tangential connection,  $PP$ . Those alone being insufficient to produce a reading order, we also introduce the four directional relations: *Up*, *Down*, *Right* and *Left*.

In order to be able to compute *hasNext* relations, we have to formalize cardinal spatial relations between panels. There are multiple ways to compute those relations between polygons. In the special case of comic books, we would like to guarantee the absence of loops within these relations which would prevent us to compute a reading order. To that end, the whole process is split in two steps according to Eq. (1) describing the *Down* relation.

$$\begin{aligned} \text{Down}(A, B) \Leftrightarrow & (DC(A, B) \wedge \forall (y_{Ai}, y_{Bi}), y_{Bi} > y_{Ai}) \\ & \vee (PO(A, B) \wedge y_{BC} > y_{AC}) \end{aligned} \quad (1)$$

If two panels are disjoint ( $DC(A, B)$ ), every point is taken into account in the calculation of the relation  $\{(x_{A1}, y_{A1}); \dots; (x_{An}, y_{An})\}, \{(x_{B1}, y_{B1}); \dots; (x_{Bm}, y_{Bm})\}$ . Otherwise, if the panels overlap each other ( $PO(A, B)$ ), their respective centroid  $\{(x_{AC}, y_{AC})\}, \{(x_{BC}, y_{BC})\}$  are used. Finally, when a panel contains another one ( $PP(A, B)$ ), we state the directional spatial relation to be non-settable. *Right* relation can be retrieved similarly by switching y-axis to x-axis. Therefore, this segmentation of the directional computation along the topological relations between two panels leads to loopless *Down* and *Right* partial orders. Besides, we can notice that  $Up = \text{Down}^{-1}$  et  $Left = \text{Right}^{-1}$ .

A reading order can now be deduced from *Down* and *Right* relations in a two-step process. Two panels being not necessarily bounded by a specific orientation relation (e.g if one contains the other), we start by computing a partial ordering (i.e. transtive, antisymmetric and reflexive relation) according to Eq. (2).

$$\begin{aligned} \text{StructuralNext}(A, B) \Leftrightarrow & \text{Down}(A, B) \vee \\ & (\neg \text{Down}(A, B) \wedge \neg \text{Down}(B, A) \wedge \text{Right}(A, B)) \end{aligned} \quad (2)$$

Then we execute a topological sorting of the *StructuralNext* relations in order to produce one total reading order  $f_1 < f_2 < \dots < f_m$ , represented by the *hasNext* relation, among the different available paths. Note that it is possible to switch from a western to an eastern reading order by simply replacing  $\text{Right}(A, B)$  by  $\text{Right}^{-1}(A, B)$  in Eq. 2.

### 3.4 Experiment

The A-Box was populated with comics material, provided by various authors such as Lamisseb [9] and Boisgibault [10], according to Fig 3. Among the thousand pages in our database, 22 have been chosen to build a ground truth on panels bounding boxes and their reading sequences. As we want to evaluate the ordering and not the accuracy of the extraction, panels have been manually segmented, producing polygons whose coordinates can be passed as entry to our algorithm. A number has been associated to each panel, matching its rank in the reading sequence. Then, the calculation of the *hasNext* relations has been processed on the basis of the segmented panels. Both data sets have been injected into our ontology, associated to two different extractors being *GroundTruth* and *Automatic*. For each *hasNext(A,B)* axiom from the *Automatic* extractor we check the existence of a matching *hasNext(A,B)* atom associated to *GroundTruth*. Polygon's sets being the same for each data sets, there cannot be any ambiguous matching between two panels from the different extractors. 338 *hasNext* relations have been automatically produced from our 22 annotated pages. Each of them has been validated by the ground truth. In order to test these relations, we have loaded our populated model into Sewelis [17]. Sewelis is a piece of software relying on Logical Information System (LIS) formalism which provides an innovative way to browse data through user-friendly query building. Applied to our data, it allows us to read our comics page by page and panel by panel.

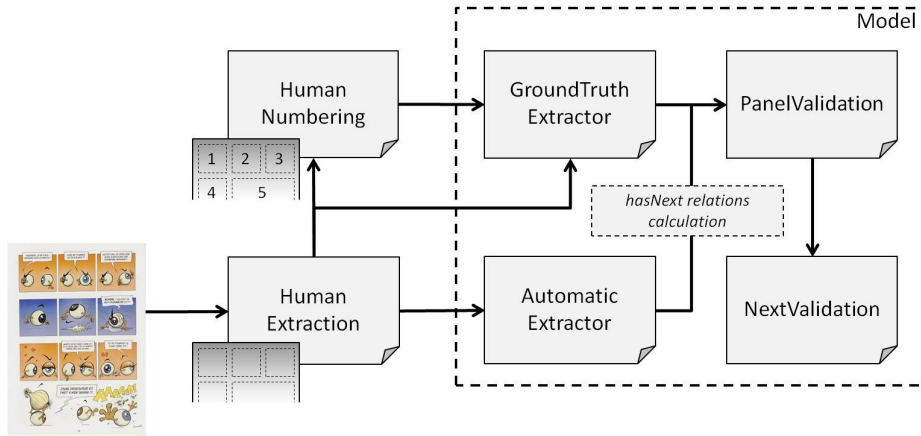


Fig. 3. The A-Box feeding macroview.

## 4 Conclusions and perspectives

This paper introduced our contribution to the semantic gap resolution in the special comic books context. We have presented an ontology allowing extracted

comic content labelling and easing the navigation by formalizing the reading order. The results can be used to ease the production of digital comics from digitized pages. Yet, it would be interesting to be able to load all different reading paths in the model as well as underline the ambiguous spots. This is definitely something that is going to be done in a very near future. Then, the process will be applied to the *text areas*, which is a slightly different problem. Our model provides a general framework to organize comic specific data and information. The next step is to make it evolve and integrate more knowledge, such as relations between characters and speech balloons or scenes identification. This will allow the development of a semantic content-based image retrieval system which would return a user, through a relevance feedback mechanism, similar images according to their semantic affinity.

## References

1. C. Rigaud, N. Tsopze, J.C. Burie, and J.M. Ogier. Extraction robuste des cases et du texte de bandes dessinées. In *Proc of CIFED'12*, 2012.
2. Takamasa Tanaka, Kenji Shoji, Fubito Toyama, and Juichi Miyamichi. Layout analysis of tree-structured scene frames in comic images. In *Proc. of the IJCAI'07*, pages 2885–2890, 2007.
3. Christophe Ponsard and Vincent Fries. Enhancing the Accessibility for All of Digital Comic Books. *eminds.uniovi.es*, I(5), 2009.
4. M. Uschold and M. King. Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 11(2):93–155, 1996.
5. T. Berners-Lee, J. Hendlers, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.
6. S. McCloud. *Understanding comics: The invisible art*. Harper Paperbacks, 1993.
7. Masashi Yamada, Rahmat Budiarto, Mamoru Endo, and Shinya Miyazaki. Comic image decomposition for reading comics on cellular phones. *IEICE Transactions*, 87-D(6):1370–1376, 2004.
8. JM. Lain and S. Delzant. *Le dessin des planches*. Eyrolles, 2007.
9. Lamisseb. *Les noels*. Editions Bac@BD, 2011.
10. P. Boisgibault, F. Marniquet, and O. Jolivet. *Boston police, l'affaire pradi*. Editions Clair De Lune, 2011.
11. Dengsheng Zhang, Md M. Islam, and Guojun Lu. A review on automatic image annotation techniques. *Pattern Recognition*, 45:346–362, 2012.
12. H. Bannour and C. Hudelot. Combinaison d'information visuelle, conceptuelle, et contextuelle pour la construction automatique de hiérarchies sémantiques adaptées à l'annotation d'images. In *Actes de RFIA'12*, pages 460–469, June 2012.
13. Bernd Neumann and Ralf Möller. On scene interpretation with description logics. *Image Vision Comput.*, 26:82–101, January 2008.
14. Anne-Marie Tousch, Stéphane Herbin, and Jean-Yves Audibert. Semantic hierarchies for image annotation: A survey. *Pattern Recognition*, 45:333–345, 2012.
15. Eunjung Han, Jonyeol Yang, HwangKyu Yang, and Keechul Jung. Automatic mobile content conversion using semantic image analysis. In *Proc. of the HCI'07*, pages 298–307, Berlin, Heidelberg, 2007. Springer-Verlag.
16. A.G. Cohn, B. Bennet, J. Gooday, and N.M. Gotts. Representing and reasoning with qualitative spatial relations about regions. In *Spatial and temporal reasoning*, pages 97–134. Kluwer, 1997.
17. Sébastien Ferré and Alice Hermann. Semantic search: reconciling expressive querying and exploratory search. In *Proceedings of the ISWC'11*, 2011.