

Automated Collage Generation – With Intent

Anna Krzeczkowska,* Jad El-Hage,† Simon Colton* and Stephen Clark†

* Department of Computing, Imperial College, London, UK.

† University of Cambridge Computer Laboratory, Cambridge, UK.

1 Introduction

One reason why software undertaking creative tasks might be perceived of as uncreative is its lack of overall purpose or intent. In computational creativity projects, while some skillful aspects may be undertaken by the computer, there is usually a person driving the process, by making decisions with regard to the intent of the artefact being produced. Such human intervention can take different forms, for instance via the supplying of background information in scientific discovery tasks, or making choices during an evolutionary art session. We are interested in whether a notion of purpose could be projected onto The Painting Fool, an automated painter we hope will one day be taken seriously as a creative artist in its own right [3, 5]. Starting with the maxim that *good art makes you think*, we have enabled The Painting Fool to produce visual art specifically to invite the viewer to interpret the pieces in the context of the world around them, i.e., to make a point about a current aspect of society. However, we do not prescribe what aspect of modern life it should depict, nor do we describe the art materials – in this case digital images – it should use. Hence, we effectively opt out of specifying a purpose for any individual piece of art. As described in section 2, the software starts with an instruction to access sources of news articles at regular intervals. Then, via text manipulation, text analysis, image retrieval, image manipulation, scene construction and non-photorealistic rendering techniques, the system produces a collage which depicts a particular news story. This is initial work, and more effort is needed to produce collages of more aesthetic and semantic value. We present some preliminary results in section 3, including some illustrative examples and feedback from viewers of the collages produced. In section 4, we describe the next stages for this project.

2 Automated Collage Generation

A schematic for the collage generation system is provided in figure 1. At regular intervals, scheduled processes begin a flow of information involving the retrieval of news articles from internet sources (with the news source specified by the scheduled job); the extraction of keywords from the news articles; the retrieval of images using the keywords; and the construction of input files for The Painting Fool. The input files specify which images to annotate and extract colour segments from, how to arrange the segments in an overall collage, and what natural media to simulate when painting the segments to produce the final piece. We provide further details of these processes below, with full details of the overall system available in [7].

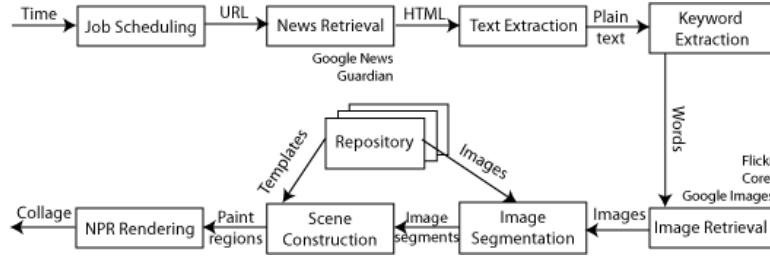


Fig. 1. Automated collage generation system overview.

- Text Retrieval and Analysis

We enabled the system to access the Guardian News website and Google News search, via their APIs. For the Guardian, the API provides access to headlines, for which there are a number of associated articles, multimedia files, blogs, forums, etc., and our system extracts the first text-based article from this list. The Google News API produces similar output from multiple news sources, from which we extract text-based news stories from the BBC, The Independent newspaper and Reuters news service. The system retrieves only English language headline articles, and we can specify whether the articles should be about World or country-specific issues only. The retrieved articles are cleaned of database information and HTML appendages to produce plain text.

Following this, we use a text analysis technique to extract a specified number of keywords from the plain text. The technique is an implementation of the TextRank algorithm [8], which is designed to extract those keywords most indicative of the content of the document. It is based on PageRank [1], the algorithm used by Google to determine the importance of a web page based on the hyperlink structure of the web. The intuition behind PageRank is that important pages will be pointed at by other important pages, a recursive notion of importance which can be assigned numerical values using a simple iterative algorithm. The intuition behind TextRank is similar: important words in the document will be *pointed at* by other important words. In the context of a document, *pointed at* is defined in [8] as being in the same context. Hence a graph representing the document can be created, where an edge exists between two words if those words appear in each other's contexts, where context is just a fixed-size window either side of the target word. PageRank uses the graph to assign numerical values to each word and extracts the most important ones. For our experiments, only nouns were extracted, as these were considered likely to be the most informative, and also the most useful keywords to use for image retrieval. Full details of the keyword extraction implementation are given in [6].

- Image Retrieval and Manipulation

The keywords extracted from the news stories are used to retrieve art materials (i.e., digital images) from the internet and local sources. The system has access to the 32,000 images from the Corel library which have been hand tagged and can be relied on for images which match the given keywords well.

We also wanted to include images retrieved from the Internet, as these add a level of surprise,¹ and a more contemporary nature to the retrieved images. We interfaced the collage generation system with both the Google images and the Flickr APIs. In the former case, the interface is fairly lightweight, given that Google supplies a set of URLs for each keyword, which point to the relevant images. In the latter case, however, a URL must be built from information retrieved from a photo-list, which is a non-trivial process. The three image sources (Corel, Google, Flickr) are queried in a random order, but when either Corel or Flickr return empty results, Google is queried, as this always supplies images. Note that we discuss experiments comparing Flickr and Google images in section 3.

- Scene Construction and Rendering

In the final stage of processing, the retrieved images are assembled as a collage in one of a number of grid-based templates. Then the system employs The Painting Fool’s non-photorealistic rendering capabilities [5] to draw/paint the collages with pencils, pastels and paints. In future, we will use the more sophisticated scene generation techniques described in [4].

3 Initial Results

The first image in figure 2 portrays a typical collage produced by the system. Here, a scheduled process happened to retrieve a Guardian news story about the war in Afghanistan, with the headline ‘Brown may send more troops to Afghanistan’. From the text, the words **afghanistan, brown, forces, troops, nato, british, speech, country, more and afghan** were extracted. Images were retrieved from Flickr accordingly, including a picture of a fighter plane, a field of graves, a young woman in an ethnic headdress and an explosion. The rendering style for this was simply to segment each image into 1000 regions and present the images in an overlapping grid with 10 slots. This example hints at the ability of the system to construct a collage which can semantically complement a news story, or even add poignancy. The second collage in figure 2 provides a hint of the possibilities for more interesting and perhaps playful juxtapositions. This was produced in response to a news story on the England versus Australia Ashes test cricket series, which had the headline: ‘England versus Australia – as it happened!’. The images of the Houses of Parliament and a kangaroo in the collage are fairly obvious additions. However, the Collage also contains a picture

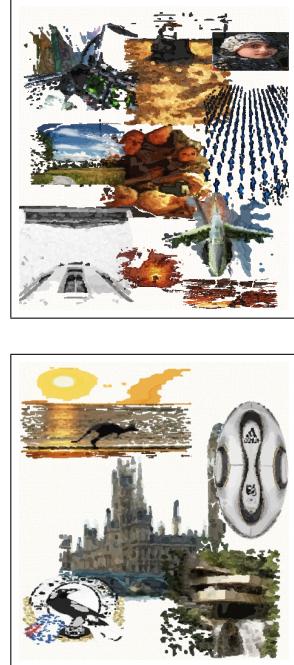


Fig. 2. Example collages produced by the system.

¹ For instance, at the time of writing, querying Flickr for images tagged with the word “Obama” returns an image of a woman body-builder as the first result.

of the famous Falling Water building built by Frank Lloyd-Wright. Upon investigation, we found that the name `wright` was extracted from the news story (as a member of the England cricket team) and the first Google image returned for that keyword is, of course, the image of Falling Water.

In order to informally assess the power of the collages to represent the text upon which they are based, we showed a collection of the collages to 11 subjects. We asked them to complete a survey where they were shown 12 news stories and 5 collages per news story, only one of which, called the *master collage* was generated from the news story (with the others generated from similar stories). To generate the 12 master collages, we varied both the number of keywords to be extracted (5 and 10) and the image source (Google and Flickr). Subjects were asked to rank the 5 collages for each news story in terms of relevance, with rank 1 indicating the most relevant and 5 the least relevant. Taking the average of the ranks, we found that the master collage was ranked as (a) the most relevant in 8 of the 12 tests and (b) the second most relevant in the other 4 tests. The most marked difference we noticed was between the collages produced with Google and with Flickr. In particular, the Google collages had an overall rank of 1.82, while the Flickr collages had an overall rank of 2.14. This highlights that image tagging in Flickr is not particularly reliable, with Google returning more relevant images. These results are encouraging, as they demonstrate that even via the abstractions of keyword extraction, image retrieval and non-photorealistic rendering, it is still possible for the collages to have semantic value with respect to the news stories from which they were derived.

4 Discussion and Further Work

The system described above is a prototype, and is presented largely as a proof of principle, rather than a finished system. We have presented an illustrative example of how the pipeline of processes can produce a collage with the potential to make viewers engage their mental faculties – in this case about warfare. The value here is not necessarily in the quality of the final artefacts – which are currently a little naïve – but rather in the fact that we had little idea of what would be produced. We argue that, as we did not know what the content of the collages would be, it cannot have been us who provided the intention for them. Given that these collages are based on current events and do have the potential to engage audiences, we can argue that the software provided the intent in this case (perhaps subject to further discussion of *intent* and *purpose* in art).

In [3], we argue that the perception of how a computational system produces artefacts is as important as the final product. Hence, the fact that the system supplies its own purpose in automated art generation may add extra value to the artworks produced. Having said that, there are a number of improvements to the process we intend to make in order to increase the visual and semantic appeal of the collages. In particular, we plan to make better use of The Painting Fool’s scene construction abilities, and to implement scene construction techniques which are aware of the context of the news story being portrayed. For instance, if the text of the news article has a distinctive plot-line, then a linear

collage might best portray the narrative of the story, with images juxtaposed in an appropriate order. However, if the major aspect of an article is the mood of the piece, then possibly a more abstract collage might best portray this. We also plan to involve text summarisation software to provide titles, wall text and other written materials for the collages. We hope to show that by stepping back from certain creative responsibilities (described as “climbing the meta-mountain” in [4]), such as specifying the intent for a piece of art, we make it possible to project more creativity onto the collage generation system than if there was a person guiding the process. Our long term goal for The Painting Fool is for it to be accepted as a creative artist in its own right. Being able to operate on a conceptual level is essential for the development of The Painting Fool, hence we will pursue further interactions with text analysis and generation systems in the future.

We would like to thank the anonymous reviewers for their useful advice. One reviewer stated that in some scientific theory formation systems, the software is not perceived as uncreative because of a lack of intent. As the engineers of scientific discovery software [2], when running sessions, we always provide intent through our choice of background material and our choices for evaluating the theory constituents. Hence, a critic could potentially argue that the software is not being creative, as it has no purpose of its own. We believe that this is true of most other scientific discovery systems, especially machine learning based approaches such as [9], where finding a classifier is the explicit user-supplied intention. The reviewer also compared the collage generation system with Feigenbaum’s famous Eliza program. We find it difficult to see the comparison, given that the collage generation system is given no stimulus from a user, whereas Eliza reacts repeatedly and explicitly to user input. A more accurate analogy in the visual arts would be image filtering, where an altered version of the user’s stimulus is presented back to them for consideration. It is clear that the notion of intent in software causes healthy disagreements, and perhaps our main contribution here is to have started a fruitful discussion on this topic.

References

1. S Brin and L Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30:1–7, 1998.
2. S Colton. Automated Theory Formation in Pure Mathematics. Springer, 2001.
3. S Colton. Creativity versus the perception of creativity in computational systems. In *Proceedings of the AAAI Spring Symposium on Creative Systems*, 2008.
4. S Colton. Experiments in constraint-based automated scene generation. In *Proceedings of the 5th International Joint Workshop on Computational Creativity*, 2008.
5. S Colton, M Valstar, and M Pantic. Emotionally aware automated portrait painting. In *Proc. of the 3rd Int. Conf. on Digital Interactive Media in Ent. & Arts*, 2008.
6. J El-Hage. Linguistic analysis for The Painting Fool. Master’s thesis, The Computer Laboratory, University of Cambridge, UK, 2009.
7. A Krzeczkowska. Automated collage generation from text. Master’s thesis, Department of Computing, Imperial College, London, UK, 2009.
8. R Mihalcea and P Tarau. TextRank: Bringing order into texts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2004.
9. S Muggleton. Inverse Entailment and Progol. *New Generation Computing* 13, 2005.