



Universidade Nova de Lisboa  
Faculty of Science and Technology  
Department of Computer Science

## **Automated Illustration of Multimedia Stories**

Diogo Delgado, N.º 27239

*Submitted in part fulfillment of the requirements for  
the degree of Master in Computer Science  
2009/2010*

Supervisor  
Prof. Dr. João Magalhães  
Co-supervisor  
Prof. Dr. Nuno Correia

28 of July of 2010



**Student n°:** 27239

**Name:** Diogo Miguel Melo Delgado

**Title:** Automated Illustration of Multimedia Stories

**Keywords:**

- Story Illustration
- Text Analysis
- Informational Retrieval
- Image-Text Association
- Vector Space Model
- User Feedback

**Palavras-chave:**

- Ilustrador de Histórias
- Análise Textual
- Recuperação de Informação
- Associação Imagem-Texto
- Modelo de Espaço Vectorial
- Feedback de Utilizador



# Resumo

---

Todos nós já tivemos o problema de ao ler um texto, esquecer o conteúdo de frases anteriores. Este problema provém da falta de atenção para com a leitura e é um problema comum nos mais jovens ou idosos. Os leitores ficam aborrecidos ou distraídos com algo mais interessante. O desafio consiste em criar sistemas de multimédia que melhorem a experiência de leitura de histórias e ajudem na retenção do seu conteúdo. A solução proposta consiste em usar imagens para ilustrar histórias, como um meio de cativar o interesse do leitor enquanto lê uma história ou fazendo o mesmo imaginar uma. Este trabalho investiga a problemática da ilustração automática de histórias. É proposta a hipótese de que um sistema automatizado de multimédia pode ajudar os utilizadores na leitura de histórias, estimulando a memória de leitura com ilustrações visualmente adequadas.

Foi implementada uma aplicação que mostra uma história e tenta capturar a atenção dos leitores, fornecendo ilustrações que fortalecem a imaginação dos mesmos. O sistema cria apresentações multimédia de notícias automaticamente (1) processando o texto da notícia frase a frase, (2) fornecendo mecanismos para seleccionar a melhor ilustração para cada frase e (3) seleccionando o conjunto de ilustrações que garantem a melhor sequência. Estes mecanismos estão enraizados em técnicas de procura de imagens e texto. Para melhorar ainda mais a atenção dos utilizadores, foi adicionada uma opção que activa a leitura oral de notícias de acordo com as suas preferências ou dificuldades de leitura. As primeiras experiências mostram que imagens do Flickr podem ilustrar artigos de notícias da BBC e proporcionar uma melhor experiência para os leitores de notícias.

Além do processo de ilustração proposto, foi também desenvolvido uma funcionalidade de personalização com o intuito de aperfeiçoar a selecção de ilustrações. Com recurso a esta característica do sistema, o utilizador pode ajudar na selecção de imagens mais relevantes.

Finalmente, no documento são descritas as avaliações empíricas realizadas a fim de testar os diversos componentes do sistema proposto e os respectivos resultados são discutidos.

---

Este trabalho foi desenvolvido no âmbito do projecto ARIA – Ambient-assisted Reading Interfaces for the Ageing-society, para a *Fundação para a Ciência e Tecnologia*, Portugal, (PTDC/EIA-EIA/105305/2008). Dois artigos foram escritos sobre o assunto e reflectem as características do trabalho. Um descreve a interface da aplicação e as suas funcionalidades e o outro descreve os algoritmos de selecção de ilustrações e as avaliações com utilizadores. Os artigos foram submetidos e aceites em duas conferências.

---

# Abstract

---

We all had the problem of forgetting about what we just read a few sentences before. This comes from the problem of attention and is more common with children and the elderly. People feel either bored or distracted by something more interesting. The challenge is how can multimedia systems assist users in reading and remembering stories? One solution is to use pictures to illustrate stories as a mean to captivate ones interest as it either tells a story or makes the viewer imagine one. This thesis researches the problem of automated story illustration as a method to increase the readers' interest and attention. We formulate the hypothesis that an automated multimedia system can help users in reading a story by stimulating their reading memory with adequate visual illustrations.

We propose a framework that tells a story and attempts to capture the readers' attention by providing illustrations that spark the readers' imagination. The framework automatically creates a multimedia presentation of the news story by (1) rendering news text in a sentence-by-sentence fashion, (2) providing mechanisms to select the best illustration for each sentence and (3) select the set of illustrations that guarantees the best sequence. These mechanisms are rooted in image and text retrieval techniques. To further improve users' attention, users may also activate a text-to-speech functionality according to their preference or reading difficulties. First experiments show how Flickr images can illustrate BBC news articles and provide a better experience to news readers.

On top of the illustration methods, a user feedback feature was implemented to perfect the illustrations selection. With this feature users can aid the framework in selecting more accurate results.

Finally, empirical evaluations were performed in order to test the user interface, image/sentence association algorithms and users' feedback functionalities. The respective results are discussed.

---

---

This work was developed in the scope of project ARIA – Ambient-assisted Reading Interfaces for the Ageing-society, for the *Fundação para a Ciência e Tecnologia*, Portugal, (PTDC/EIA-EIA/105305/2008). Two papers have been written on the subject and reflect the frameworks properties. One describes the application’s interface and features and the other describes the illustration selection algorithms and user tests performed. These papers were submitted and accepted in two conferences.

---



# Acronyms

---

Term Frequency – Inverse Document Frequency	TF-IDF
Text-to-Speech	TTS
Information Retrieval	IR
Story Sequence Consistency	SSC
Relevance Feedback	RF



# Contents

---

<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Motivation	2
1.2 Objective	3
1.3 Proposed Framework	3
1.4 Organization	4
<b>Chapter 2 Assisted Story Reading</b>	<b>5</b>
2.1 Related Work	5
2.2 Framework Overview	6
2.3 User Interface	9
2.3.1 News Selection	9
2.3.2 News Reading	11
2.3.3 News Reading Debug	12
2.4 Controller	15
2.5 Text-To-Speech	16
2.6 User Evaluation	17
2.6.1 General Assessment	18
2.7 Summary	19
<b>Chapter 3 Automatic Sentence Illustration</b>	<b>21</b>
3.1 Related Work	21
3.1.1 Text Analysis	22
3.1.2 Illustration Analysis and Selection	24
3.2 Standard Processing	25
3.3 Semantic-based Comparison	30
3.4 Story Sequence Consistency (SSC)	34
3.5 User Evaluation	35
3.5.1 News Dataset: BBC Web News	35

3.5.2	Images Dataset: Flickr Photos	36
3.5.3	Data Model	36
3.5.4	Experiment Design	38
3.5.5	Results and Discussion	39
3.6	Summary	41
<b>Chapter 4</b>	<b>User Feedback based Illustration</b>	<b>43</b>
4.1	Related Work	43
4.2	Personalized Sentence Illustrations	45
4.2.1	Explicit Feedback	45
4.2.2	Timed Feedback	47
4.3	Improved Story Sequence Consistency	49
4.4	User Evaluation	49
4.4.1	Experiment Design	50
4.4.2	Results and Discussion	52
4.5	Summary	55
<b>Chapter 5</b>	<b>Conclusions and Future Work</b>	<b>57</b>
5.1	Conclusions and Contributions	57
5.2	Future Work	59
<b>References</b>		<b>61</b>
<b>Annexes</b>		<b>67</b>
I.	Assisted news reading with automated illustration	69
II.	Automated illustration of news stories	73
III.	Story illustrator & Illustration Methods Assessment Questionnaire	79
IV.	User Feedback Assessment Questionnaire	81

# Figures

---

Figure 1.1: Assisted reading.	2
Figure 1.2: Proposed framework for automated story illustration.	4
Figure 2.1: “Ken Burns” effect.	6
Figure 2.2: System architecture diagram.	7
Figure 2.3: Automated Story Illustrator class diagram.	8
Figure 2.4: Splash screen.	9
Figure 2.5: News Subject selection.	10
Figure 2.6: News Section selection.	10
Figure 2.7: News Headline selection.	10
Figure 2.8: News Reading mode.	11
Figure 2.9: News Reading Debug mode.	12
Figure 2.10: News Reading mode with comments.	13
Figure 2.11: News Reading Debug mode with comments.	14
Figure 2.12: Sequence diagram describing the process of news reading.	15
Figure 3.1: Text processing and tf-idf weighting applied on a sentence.	28
Figure 3.2: Text processing and tf-idf weighting applied on a picture.	28
Figure 3.3: WordNet browser.	31
Figure 3.4: Sliding Window on Story Illustration application.	34
Figure 3.5: User assessment of the illustration methods.	39
Figure 4.1: News Reading mode (highlight in the feedback functionalities).	46
Figure 4.2: Relevance feedback example of a user reading speed.	48
Figure 4.3: News Headline with highlight on the feedback options.	52
Figure 4.4: General user assessment of the feedback methods.	53
Figure 4.5: Results for user feedback methods per news.	53

Figure 4.6: Assessment of the feedback methods divided by news read.	54
Figure 4.7: Feedback rates on the user evaluation.	54

# Tables

---

Table 2.1: Test users' age.	17
Table 2.2: Users' general assessment of the system.	18
Table 2.3: Users' general assessment of the system (continuation).	19
Table 2.4: User visual memory.	19
Table 3.1: Illustration of "Online maps 'wiping out history'".	29
Table 3.2: Illustrations of "UK escapes the worst case of bluetongue".	30
Table 3.3: Illustrations generated by the different illustration methods.	33
Table 3.4: News articles table.	37
Table 3.5: Pictures table.	37
Table 3.6: News words table.	37
Table 3.7: Test users' age.	38
Table 3.8: User assessment of the illustration methods.	39
Table 4.1: Test users' age.	50
Table 4.2: Distribution of feedback methods per news and user.	51
Table 4.3: General user assessment of the feedback methods.	52





## Chapter 1

# Introduction

When reading a book it is common for readers to create images in their minds that portray the text. For example, in Mary Shelly's *Frankenstein*, the monsters image, depicted in our heads, may differ from person to person. Sometimes, it is tempting to spark the readers' imagination with book illustrations as it better engages the reader in the story. Over the years the possibility of using computers to choose the best illustrative pictures has been studied.

One common use of pictures is to share stories, experiences and travels with family and friends. Everyone has browsed a photo album while the photographer describes the story behind the pictures. The World Wide Web facilitates the sharing of digital photos and has fostered "digital storytelling" using photo collections. Nowadays there is no need to send photographs by e-mail since it is possible to upload the pictures to a social network website like *Facebook* or *Flickr* providing access to everyone involved in the album, and more if it is the uploaders' desire. The story is told with a set of photographs and a few comments added to the pictures – the human mind fills in the gaps. This is well discussed by Speer, Reynolds et al. (2009) that registered human brain activity while reading text stories. Reading about different visual scenes and motor experiences would activate particular zones of the brain.

In this setting the Web supplies us with millions of tagged images, which can be useful to illustrate stories with real images that capture readers' curiosity and imagination. Searching for a video on YouTube or a photograph on Flickr requires

users to describe the most important aspects of what they wish to find. It is not common to use full sentences as “Video where Mary went to a night festival in London”, but words like “Mary”, “festival” or “London”. These words are addressed as *keywords*. They define the sentence content and make search engines more effective. In our work, text keywords are used to search picture *tags* – user supplied text annotations.

## 1.1 Motivation

Many areas exist where the automatic creation of multimedia presentations from text is of great value, e.g. entertainment, media, journalism, children’s picture books. For example, when journalists write news, they need to select images to illustrate text parts of the whole news article. These illustrations are usually selected from a repository or from specific photos taken at the news setting.



Figure 1.1: Assisted reading.

Automated story illustration may also be of use when in need for attention enhancement. Pictures often improve the readers’ attention: Mayers et al., (1995), recorded an experiment where students learning a subject from an only-text explanation performed worse from students who learned from illustrated text explanations. Cianciolo<sup>1</sup> explained the value of picture storybooks in general to children's intellectual development:

*It is the content of the picture book, and the quality of the literary and artistic amalgam resulting from the combination of words and illustrations used to express that content, which facilitate the development of the reader's*

---

<sup>1</sup> P. J. Cianciolo, Picture Books for Children

*imaginative, creative and critical thinking, satisfying his or her basic need to know, or fulfilling an innate craving for the beautiful.*

## **1.2 Objective**

The goal of this thesis is to create a framework to assist news readers by generating an audio/visual presentation of the news to improve their attention. The presentation is generated with fractions of the original text with the appropriate image illustration framed in the context of the text. The key aspect of our contribution is a method to perform a semantic comparison between a range of sentence terms and image tags. This challenge is tackled in three ways. First, the system expands sentences with a linguistic ontology to enclose all possible linguistic meanings. Second, because reading is a sequential process, the system guarantees the sequential consistency of the illustrations by selecting a set of semantically coherent images. Third, the system incorporates a user feedback feature to consider user preferences.

## **1.3 Proposed Framework**

In this thesis the proposed framework architecture, depicted in Figure 1.2, has two main building blocks:

- **Illustration Selection** – Module where text processing techniques and text metrics are used to quantify the relation between a text terms and image tags. User tests will determine the best illustration method.
- **User Feedback** – Module responsible for collecting feedback from the reader. User tests were performed to realize the preferences regarding the feedback algorithms. This extra information allows the system to deliver an advanced user experience.

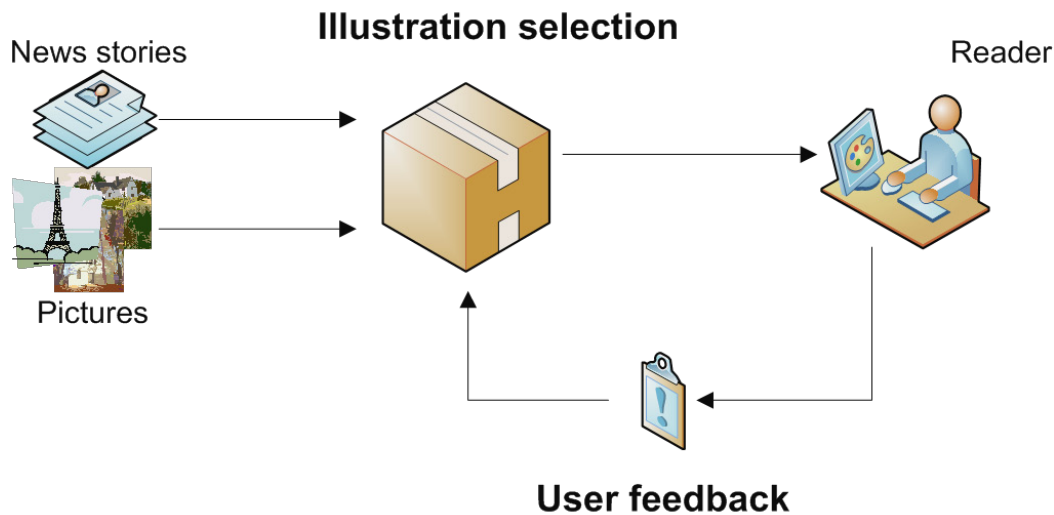


Figure 1.2: Proposed framework for automated story illustration.

## 1.4 Organization

This thesis is organized as follows:

- **Chapter 2** describes the fundamentals of story illustration and presents the proposed framework architecture. The user interface and the systems features are carefully detailed and evaluated by real users in a controlled experiment;
- **Chapter 3** examines the systems core algorithms: some solution are proposed for selecting illustrations framed in the text's content. A discussion of related text processing algorithms, namely term weighting, text summarization, keyword extraction, and automated story illustration is presented. The solutions are detailed and evaluated by real users in a controlled experiments to assess the algorithms' accuracy;
- **Chapter 4** presents the user feedback features of the framework. The basics of feedback are presented and an algorithm to measure the user's preferences is proposed. Also, the feedback features are evaluated by real users;
- **Chapter 5** presents the conclusions and prospects for future works;

## **Chapter 2**

# **Assisted Story Reading**

An application was implemented to research automated story illustration as a method to increase the readers' interest and attention. This application is a framework to develop ideas and research algorithms suited to automated story illustration. In this chapter the application prototype is described with special focus on the user interface. The system architecture is presented and a formal description of the developed application process, a UML model diagram, was included with the appropriate descriptions.

### **2.1 Related Work**

The creation of multimedia stories by humans is a process that relies on the authors' imagination and on the media resources they produce and edit. Currently it is common to use computer software for this task as is the case of the work proposed by Balabanovic et al. (2000). They describe the implementation of a device that provides a convenient way of sharing digital photographs and associated stories with family and friends, in an easier way than the conventional album browsing. While Balabanovic proposed a system to help humans assemble a multimedia story, in this thesis we aim at researching a multimedia system that replace humans in this task by taking a news story and linking meaningful image illustrations to each news segment.

Automatic text-to-scene conversion using computer graphics techniques has been studied for several years and numerous papers have been published on this topic. One of such examples is the StoryRoom (Alborzi, Druin et al. 2000), a

physical interactive spaces for children created at the University of Maryland. Other projects have focused on understanding the text: WordsEye system developed at AT&T Labs (Coyné and Sproat 2001) is a system that parses natural language and converts it into three-dimensional scenes that represent the given text. The goals of the former and the latter are similar.

Story illustration has been approached in other areas. In the movie theater industry, film-makers Kevin Macdonald and Sir Ridley Scott are to make a film based on videos submitted by YouTube users. Following a script, the directors can choose videos that best suit the scenes. YouTube users will be able record their daily life in 24 July and upload their videos for the project. It is an idea that opens the possibility of a user-generated film. In video production, one technique used for illustrating stories and events is popularly known as the “Ken Burns” technique. The technique is principally used in historical documentaries where film or video material is not available. Life is given to still photographs by slowly zooming in on subjects of interest and panning from one subject to another while the narrator speaks. Figure 2.1 shows example where an image is created from a wider image by simply zooming in. An illusion of motion can be created by panning through the original image, keeping the viewer visually entertained.

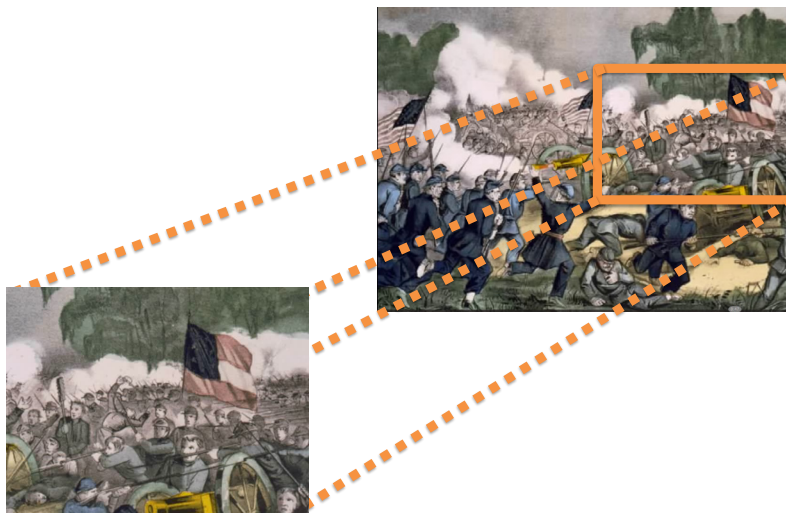


Figure 2.1: “Ken Burns” effect.

## 2.2 Framework Overview

The framework supports single-users and it has been developed as a Microsoft Windows application in C++ programming language. Its architecture

follows the traditional Model-View-Controller pattern with the data storage corresponds to the model, the user interfaces correspond to the view and the controller implements the logic behind the story telling functionality and the algorithms to rank candidate illustrations.

The assisted news reading application is composed by three dialog-based windows (News Subject, News Section and News Headline) and a full screen mode (the News Reading mode). The actual display of the news occurs in the news reading screen. The time to display a sentence is calculated based on the number of words. The key aspects of the system available to users are a text-to-speech functionality, aimed at users who have eyesight problems or reading difficulties, and most importantly, an automated illustration method to illustrate news articles. This last method is based on the relationship between image tags and a range of sentences (the main sentence and its neighboring sentences). Both news and picture information are stored in a database for an easier access to both the text data and the image tags. For the purpose of inserting news text and pictures tags in the database, two applications were implemented. They pre-process the text by running through both datasets of news articles and pictures, collecting the needed information and computing the news terms and image tag weights. The database only makes available a post-processed version of the data to the story illustration prototype. The data processing tasks performed by these applications are described in sections 3.5.1 and 3.5.2.

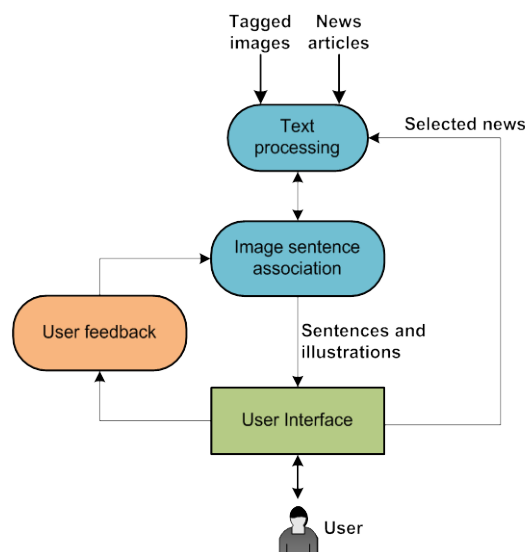


Figure 2.2: System architecture diagram.

For a full understanding of the implementation, Figure 2.3 offers a clear view of the application's class diagram showing the most important classes, attributes and implemented methods.

The main class is the CIllustratorApp and it has a private member of the class Controller. Each interface class is initialized with the main class as a parent for a straightforward change between windows. Since there were many news articles in the used dataset, we chose to divide the news in categories (sections) and divide the sections in subjects, therefore the usage of the two dialog boxes previous to the actual news reading mode.

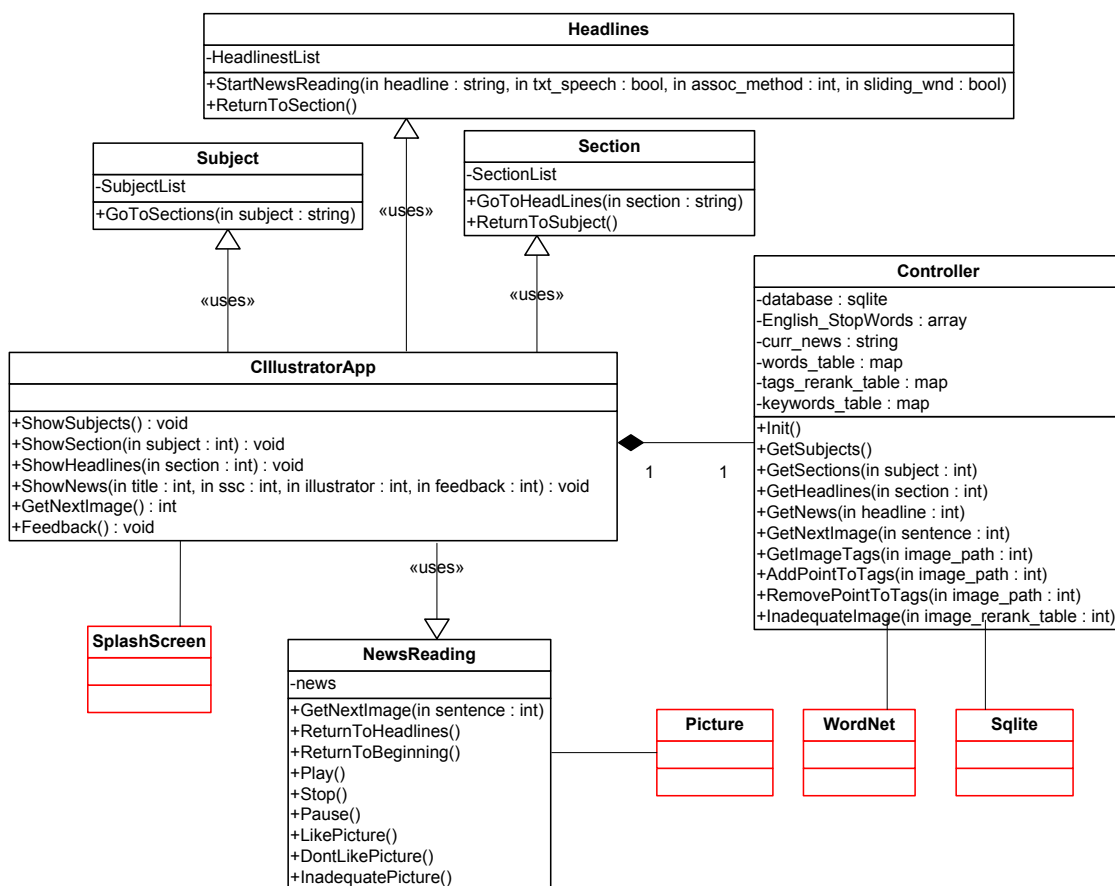


Figure 2.3: Automated Story Illustrator class diagram.

In the class diagram, the principal classes and methods are represented while classes in red represent open source classes used in the framework. These auxiliary classes were used in the application's development and therefore are part of the class diagram. Since these classes were already implemented and only few of their methods were used, their descriptions are not provided in the diagram. Also, since it



is not available to the user and therefore not part of the final application, the class News Reading Debug is not presented in the diagram (this mode is described on section 2.3.3). Initially, in order to be able to perceive the correctness of the associations' algorithms while displaying the text, a debugging interface was implemented. This user interface differs from the news reading mode because it displays not only news sentences and pictures, but also displays image tags and their contribution to the image-sentence association computation. This mode is available only to development users by pressing the right combinations of keys.

To choose the best illustration for a sentence, all image-related text data is needed. For this reason all image-related data is loaded into memory at start-up. News data is only loaded when necessary, which is after the news and illustration method are selected.

## 2.3 User Interface

In this section, the user interface of the prototype is described. We carefully detail the application interface and explain what data is presented to the user.

At application start (and during loading times), a splash screen (Figure 2.4) shows the loading progress and afterwards, the articles subjects are presented. The splash screen was added for the user to be able to perceive that the application is running.

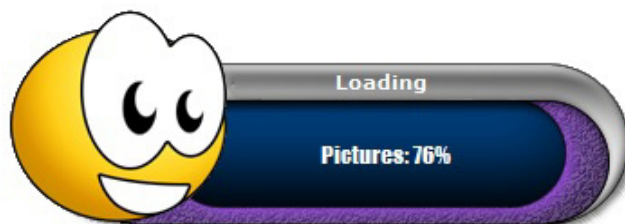


Figure 2.4: Splash screen.

### 2.3.1 News Selection

When the application starts, the *News Subject* dialog displays the existing subjects of news articles. Users start by selecting the subject they intend to read about – dialog box on Figure 2.5. Once users have selected a subject, the next button will take them to the next dialog box. On the second dialog box, Figure 2.6, users

select the section of the news articles. The next button will take them to headlines selection dialog box illustrated on Figure 2.7

The *News Headline* selection dialog box appears after selecting the news category or by pressing the *back* button in the news reading window. It displays the headlines of all the news of the previously selected section. Users select not only the news headline but also the news illustration method, combining it with a memory-based method and text-to-speech functionalities. Once users finish this last dialog box, the next button starts the presentation of the news article.

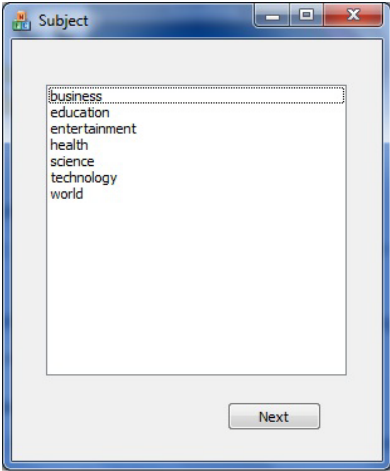


Figure 2.5: News Subject selection.

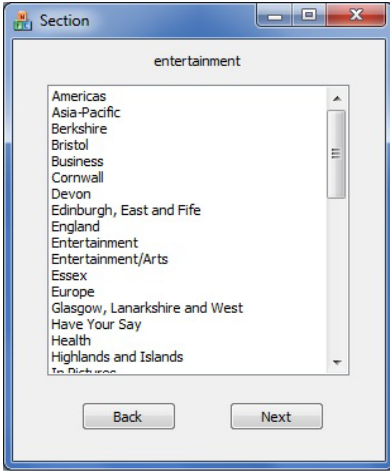


Figure 2.6: News Section selection.

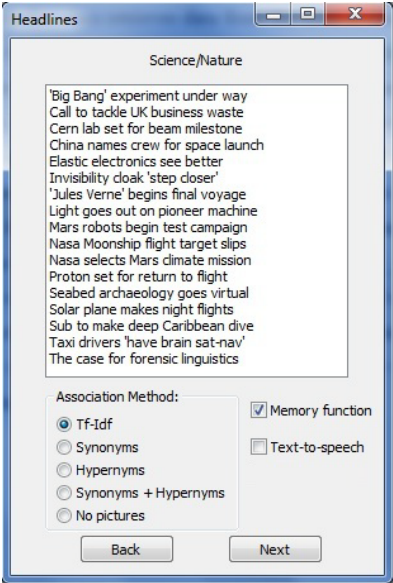


Figure 2.7: News Headline selection.

### 2.3.2 News Reading

The news article is rendered in full screen mode as illustrated by Figure 2.8. News text is presented in a sentence by sentence fashion and each sentence is illustrated by automatically selected images. The news title is presented at the top of the screen and the current sentence is presented below the image. During the news presentation, illustrations are passed as a slide show with captions corresponding to sentences of the current news article. To assist users, we provided a text-to-speech functionality to read the text out-loud. On the full screen view, the user can pause, resume or stop the news presentation (*Pause*, *Play* and *Stop* buttons respectively). The news presentation can be played faster with the *Next* button or slower with the *Previous* button by jumping between sentences.

Users can provide positive feedback with the *Like* button and negative feedback with the *Don't Like* button. Feedback will be applied to tags associated to the currently being displayed image. The *Inadequate* button serves to penalize an image and prevent the application from selecting that same image for that particular news article. The details of the user feedback functionalities are presented on Chapter 4.

Finally, users can return to the first dialog (*Home* button) and select a different subject, section and headline or go to the headlines dialog (*Back* button) to select an article from the same section or a different illustration method.



Figure 2.8: News Reading mode.

### 2.3.3 News Reading Debug

For “debugging” purposes, a different reading mode was created, where the three most relevant image tags are displayed to show which terms (or term synonyms and term hypernyms) were the major contributors to select the current image. Also for the same purpose, two additional images are displayed on the window with the respective top three tags. These images represent the three images with the highest ranking for illustrating the current sentence. For the user version, this mode is not available. It was only created with the purpose of helping the developers understand if the illustration algorithms were correctly associating text terms with image tags (or synonyms and hypernyms).

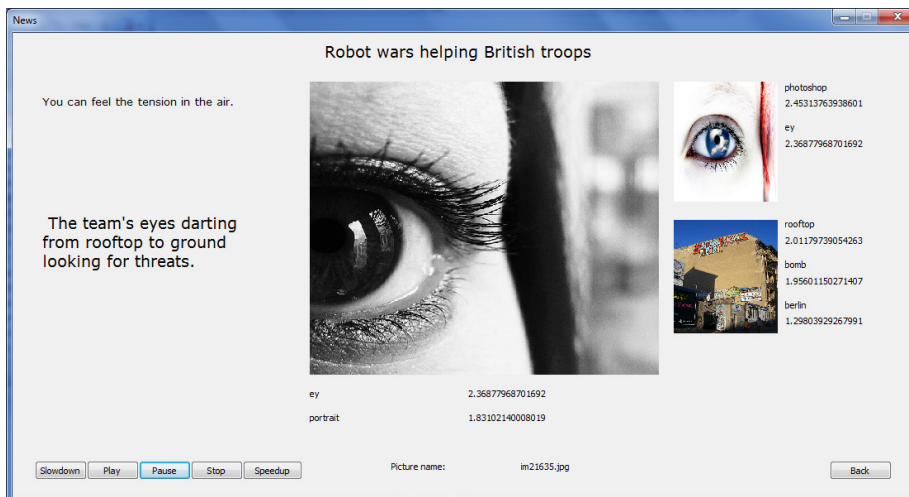


Figure 2.9: News Reading Debug mode.



Figure 2.10: News Reading mode with comments.

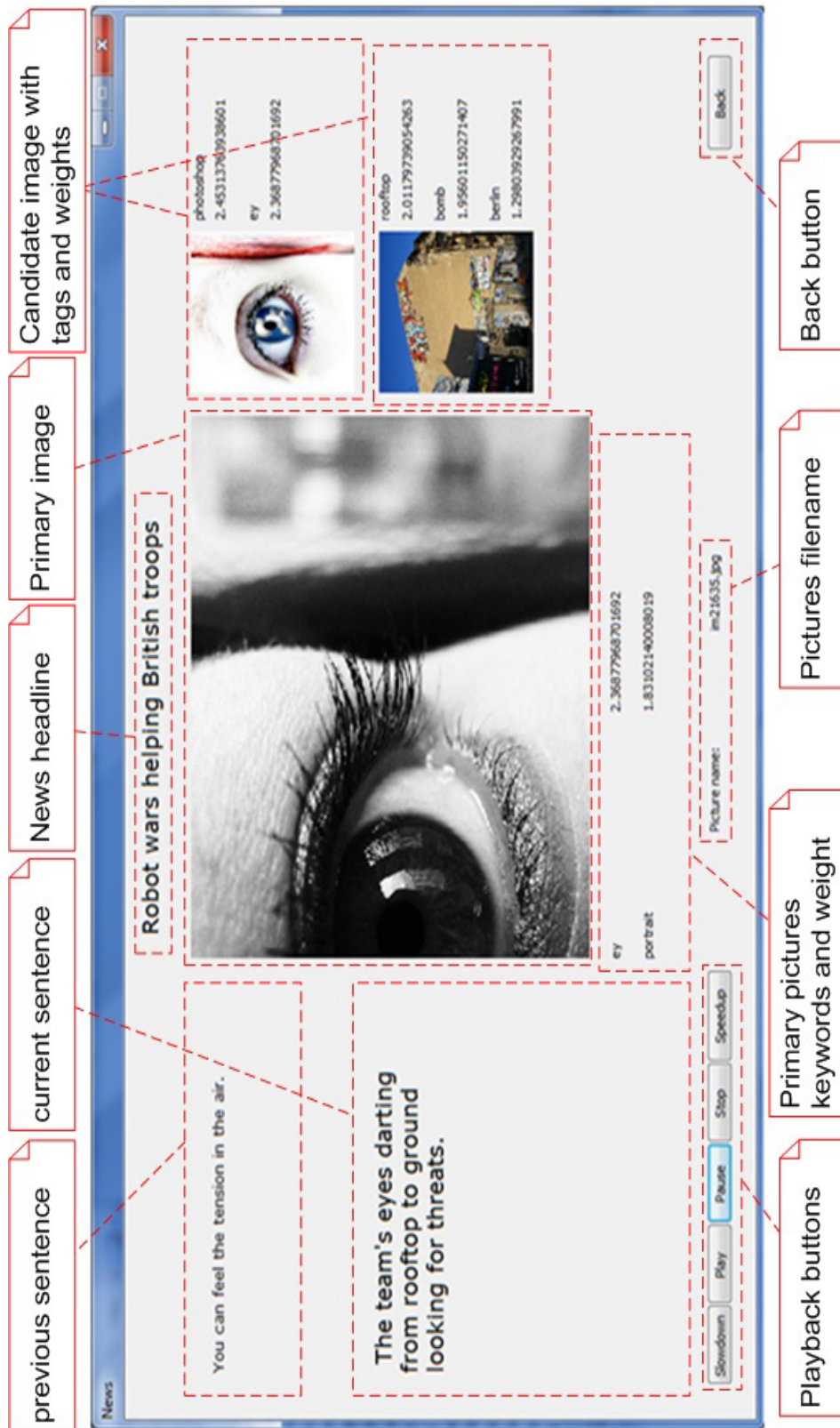


Figure 2.11: News Reading Debug mode with comments.

## 2.4 Controller

The controller class retrieves data from the database and loads the tables into memory. The Controller class controls the whole data exchange from news text to choosing the pictures to associate to a given sentence or paragraph. The class performs the calculations concerning picture ranks and the attribution of new tag weights provided by the user feedback. Since searching lists of strings is a slower process, for an application faster access and processing, the database tables ID's are used in the computations. The association algorithm uses variables of type *int* (that represent the ID's) to represent terms and tags.

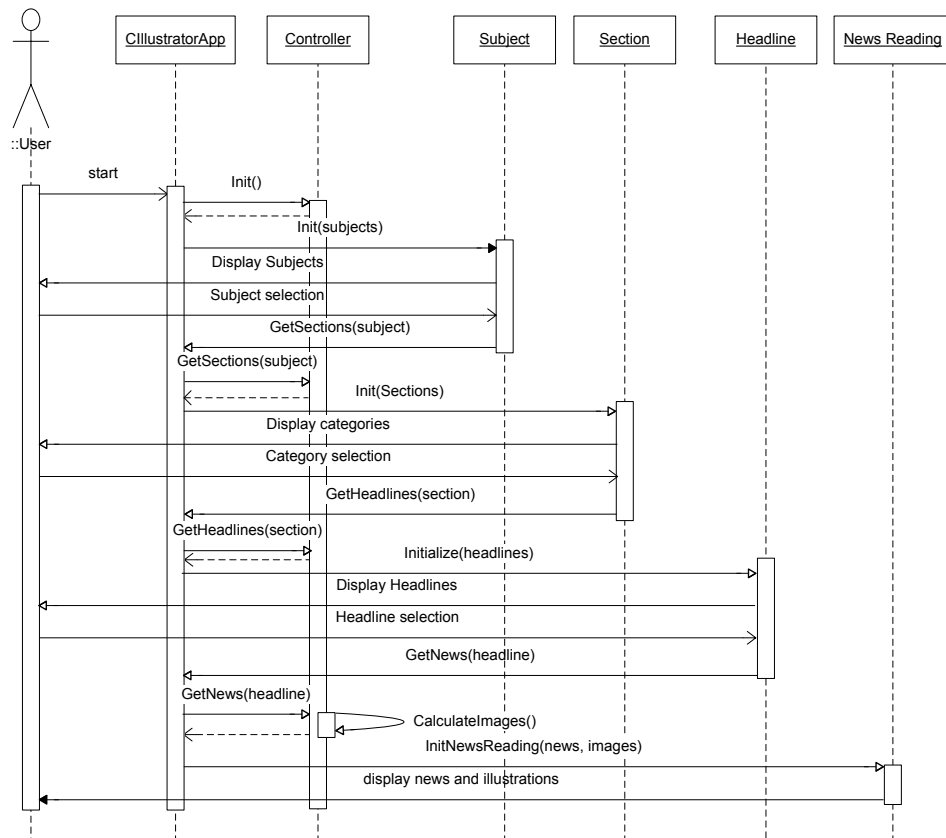


Figure 2.12: Sequence diagram describing the process of news reading.

The prototype was implemented with this characteristic to simplify all other classes in the framework. The controller is the core of the application as it is responsible for the computation of the illustration algorithm and the processing of the data provided by user feedback.

Next, a sequence diagram details the classes involved and the flow of events in the process of displaying news. This diagram concerns a standard flow of events, because in an alternative use of the application, the user can interact with the program by using the feedback functionalities.

Every dialog class communicates only with the main application class. They have to go through it to receive the needed data. One could think that such a centralized message passing mechanism is not an efficient way to forward information (the communication could be done directly between interface classes and Controller class) but we chose this approach to reduce the effort of adding new functionalities. For example, the introduction of new dialog boxes would just require the implementation of an interface integrated with the main application class. Thus, new classes only have to adapt to the main class methods and this will forward requests to the Controller.

## 2.5 Text-To-Speech

The field of text-to-speech (TTS) conversion has seen a great increase in both research community and commercial applications over the past decade. Formerly, to integrate TTS into an application, developers had to search the engine providers, select one from the available choices, buy a copy of the software, and install it. The installation, maintenance and customization of a TTS engine can be a tedious process. Chu et al. (2007) proposes a service that consists of a simple, easy-to-use platform that enables users to voice-empower their content, such as podcasts or voice greeting cards. The target users of the service include Web-based service providers such as voice greeting card companies, as well as numerous individual users who regularly or occasionally create voice content such as Podcasts or photo annotations. New and enhanced services in the telecommunications industry have increased interest in speech I/O for the workstation. In Wei et al. (2004), face tracking and lip reading were combined in a system to facilitate human computer interaction (HCI) applications in speech learning.

Some of the most popular speech recognition systems are Sphinx4<sup>2</sup> or Microsoft Speech API<sup>3</sup>. Sphinx4 is an implementation of a speech recognition

---

<sup>2</sup> <http://cmusphinx.sourceforge.net/sphinx4/>



system written entirely in Java programming language. It was created via a joint collaboration between the Sphinx group at Carnegie Mellon University, and several other renowned brands – Sun Microsystems Laboratories, Mitsubishi Electric Research Labs (MERL), and Hewlett Packard (HP), University of California at Santa Cruz (UCSC) and the Massachusetts Institute of Technology (MIT). The Microsoft Speech API, commonly known as Microsoft SAPI, is a commercial speech recognition and speech synthesis engine. The speech recognition engine included with SAPI provides a high-level interface between an application and speech engines.

Given that the objective of this work is to produce a framework that improves attention when reading, a feature of TTS was added to read news orally. Since the application was implemented in C++ programming language, the TTS feature was included based on the Microsoft speech SDK.

The implemented prototype was prepared to receive text only in English language. Additional language integrations would also require new languages in the TTS element. TTS is another feature that attempts to draw the users' attention to the news at hand and was evaluated, as part of the framework, by real users.

## 2.6 User Evaluation

To assess the general approach we conducted a user evaluation with twenty-three subjects (five females and seventeen males). The subjects were computer science academics and graduate students. Table 2.1 presents the age demographics of the group of users. The user study was performed on a Windows machine and all subjects had headphones to isolate ambient noise. Details concerning news and image data and their processing are given on sections 3.5.1 and 3.5.2.

Age	
Min	22
Mean	25
Median	27.22
Max	43

Table 2.1: Test users' age.

---

<sup>3</sup> <http://www.microsoft.com/speech/technology.aspx>

Subjects read six news articles each and after reading the news they were asked to identify the application’s most valuable features. Subjects were asked to quantify their opinion about the features with values between 1 (worst) and 5 (best). After reading the news articles, subjects were shown an image, previously used in the test to illustrate one sentence from a news article, and were asked to associate the illustration with the correct news – to assess the users’ memory. In the end, users were asked to give comments regarding the application and some gave suggestions used in the next versions of the framework.

**2.6.1 General Assessment**

The general assessment conducted at the end of the experiment, measured the effectiveness and viability of the idea.

Results in Table 2.2 suggest the application can indeed assist users in reading news. Every user preferred reading with the illustrations but, only 72.2% preferred audio. A few users did not like the audio as much as the visual content and some of the reasons were that the narrators’ voice was too “robotic” and unnatural.

Question	Result (yes)
Do you prefer multimedia presentation of news?	100,00%
Do you prefer news with sound?	72,22%

**Table 2.2: Users’ general assessment of the system.**

The last three questions were rated from 1 (worst) to 5 (best) to assess if users found the illustrations to be useful for imagining the story, enjoyable for making the news reading more interesting, or if the selected illustration were adequate or not. This last question provided us with an interesting result and feedback: users found some pictures to be related to the news content but not adequate because of the news tone. For example, if the news content was sad or tragic, then an image portraying happiness or some positive emotion would not be adequate for that specific news article.

Question	Result
Were the illustrations useful?	3,83
Were the illustrations enjoyable?	3,83
Were the illustrations adequate?	3,50

**Table 2.3: Users' general assessment of the system (continuation).**

In the question where users connect a picture with an article, the large majority answered correctly which suggests that the illustrations were useful. Table 2.4 presents the results concerning users' visual memory. Users were shown an image and asked if they recalled the corresponding news. Results show that even after watching 6 news articles with more than 10 images per article, the majority of users were still capable of correctly recalling the news general content from one single image. Few subjects did not remember the article while some recalled the articles' title but could not fully remember its content. We believe this provides a good indication on how subjects' memory performs in presence of a visual stimulus.

Recalls news title/content?	Result
No	8,70%
Vaguely	17,39%
Yes	73,91%

**Table 2.4: User visual memory.**

## 2.7 Summary

In this chapter, the issue of assisted reading was introduced resulting in an application for displaying stories with illustrations. The framework's architecture and process of text illustration were described. A database was created with the purpose of storing pre-processed data concerning images and news articles. The controller class calculates the similarity between images and news sentences and provides the User Interface with illustrations for the news presentation. News articles are read at the News Reading mode, which is a full screen window with playback options. Furthermore, a TTS feature was introduced, to provide assistance to the user. Finally, user tests were performed in order to determine the frameworks usability.

Following this chapter, a demo paper was published describing the framework objective, architecture and user evaluation as presented in this chapter. This is available as annex:

- Diogo Delgado, João Magalhães, Nuno Correia, “Assisted news reading with automated illustration”, ACM Multimedia, technical demo, Florence, Italy, October 2010.

## **Chapter 3**

# **Automatic Sentence Illustration**

The application proposed in this work computes the degree of association between news text and image tags to choose the image that best illustrates the text context. The selection process is divided in three steps: first we perform standard text processing techniques such as sentence extraction, stop-words removal and stemming – the stemming process is also applied to image tags; second, we weight each image annotation and news keyword with a weighting technique to quantify its importance in the news/image collections; and thirdly, we compute the similarity between weighted image tags and weighted news terms to determine the image most suited to illustrate each news sentence. This baseline approach is then extended with an ontology based expansion and a method to improve the coherence of sequential illustrations selections. Also in this chapter, user tests are presented and discussed.

### **3.1 Related Work**

Story illustration has a rich source of prior work and has links to research areas such as Image Processing, Information Retrieval and Natural Language Processing. This section describes some of the work related to this subject and represents a portion of the full list of works. Story illustration addresses the problem of finding the best set of pictures to describe a piece of text. On the other hand, linguistic indexing of pictures attempts to find the most suitable text to retrieve a given image. In the context of automated story illustration, our framework selects images based on their annotations to illustrate a given text.

### 3.1.1 Text Analysis

Automatic text summarization is an active research field where the objective is to extract content from a textual document and present the most important aspects to the user in a condensed form and in a manner sensitive to user or applications needs (Mani 1999; Berger and Mittal 2000). This technique can be of use for story illustration, given that it simplifies and organizes sentences or paragraphs for word comparison algorithms.

In the information retrieval research field, term weighting is an important technique for document retrieval, search engines, document summarization, text mining, and more. By computing the importance of each text term, we can easily find the document to read or learn the relation among documents. A popular term weighting technique is the *tf-idf* heuristic, which computes high values for terms that appear frequently in a document, but rarely in the remainder of the corpus. Luhn (1957) was pioneer in studies that determine the terms weight by their occurrence. Since then, other measures of term occurrences have been developed (Noreault, McGill et al. 1981; Jones 1988). Kageura and Umino (1996) summarized five groups of weighting measure: (i) a word which appears in a document is likely to be an index term; (ii) a word which appears frequently in a document is likely to be an index term; (iii) a word which appears only in a limited number of documents is likely to be an index term for these documents; (iv) a word which appears relatively more frequently in a document than in the whole database is likely to be an index term for that document; (v) a word which shows a specific distributional characteristic in the database is likely to be an index term for the database.

In some papers, the co-occurrence of two terms in a sentence is counted. Co-occurrence has long attracted interest in computational linguistics. Pereira, Tishby et al. (1993) clustered terms according to their distribution in particular syntactic contexts. From a linguistic point of view, Tanaka and Iwasaki (1996) uses co-occurrence matrices of two languages to translate an ambiguous term. On the subject of probabilities, Dagan, Lee et al. (1999) describes a method for estimating probability of previously unseen word combinations using available information on “most similar” words. When there is a need for keyword extraction in a single document, other algorithms may be more effective. Matsuo and Ishizuka (2004) describes an algorithm based on co-occurrence statistical information, in which co-

occurrences between terms are counted and the degree of bias of its distribution is measured by the  $\chi^2$  measure.

The aim of keyword assignment is to find a small set of terms that describes a specific document, independently of the domain it belongs to. An approach on automatic extraction of keywords using a supervised machine learning algorithm was proposed by Hulth (2003). Hulth came to the conclusion that by adding linguistic knowledge to the representation rather than relying only on statistics returns improved results when compared to professional keywords indexers. Since our framework makes use of a large corpus, we prefer to use the proved methods for term weighting.

Finding potential terms when no machine learning is involved in the process and by means of part-of-speech (PoS) – process of marking the words in a text (corpus) as corresponding to a particular part of speech, based on both its definition, as well as its context – patterns is a common approach. For example, Barker and Cornacchia (2000) discuss an algorithm where the number of words and the frequency of a noun phrase, as well as the frequency of the head noun is used to determine what terms are keywords. An extraction system called LinkIT (Evans, Klavans et al. 2000) compiles the phrases having a noun as the head, and then ranks these according to the heads' frequency. This is similar to our combination of the stemming algorithm (Porter 1980) and the *tf-idf* method, but in which we apply not only to nouns but to every “non-stopword”.

Boguraev and Kennedy (1999) extract technical terms based on the noun phrase patterns suggested by Justeson and Katz (1995); these terms are then the basis for a headline-like characterization of a document. The study performed by Daille, Gaussier et al. (1994) applied statistical filters on extracted noun phrases and concluded that term frequency is the best filter candidate of the scores investigated. When PoS patterns are used to extract potential terms, the problem lies in how to restrict the number of terms and only keep the ones that are relevant because some words are ambiguous and can represent more than one part of speech at different times.

### 3.1.2 Illustration Analysis and Selection

In the area of information retrieval, the story illustration problem has been approached as an image ranking and selection problem: “how to choose the best set of images from an image database to illustrate a piece of text”. Several efficient image retrieval systems have appeared in the last decade (Ma and Manjunath 1999; Smeulders, Worring et al. 2000; Wang, Li et al. 2001; Carson, Belongie et al. 2002). Some focus on quantifying image similarity – returning images similar to one image that serves as query. For Barnard and Forsyth (2001) the idea of auto-illustration as an inverse problem is introduced. Statistical associations between images and text were used to find images with high likelihood, given a piece of text.

Computing the degree of association between blocks of information is a difficult task – mutual reinforcement principle-based methods have been widely reported in the computation of these links and have been successfully used in quantitative literacy, data analysis, journal evaluation and Web search (Brin and Page 1998; Kleinberg 1999; Li, Shang et al. 2002; Joshi, Wang et al. 2006). Joshi, Wang et al. (2006) proposed a technique for image ranking and selection based on a mutual reinforcement principle and a discrete Markov chain model. The set of candidate images is assumed to form a graph with the images acting as nodes and image similarities forming the weights of the edges. Under a special condition, the image selection can be modeled as a random walk in the graph. The use of link analysis techniques where the underlying structure is a graph has been widely reported. Google's *pagerank*, measures the importance a particular web-page based on the number of links with other pages (Brin and Page 1998). Kleinberg (1999) attempts to find *hubs*, Web pages pointing to many important sites, and *authorities*, important Web sites pointed to by many other pages. In Li, Shang et al. (2002), improvements have been made on the HITS algorithm, where weights have been assigned to each link, based on textual similarities between Web pages.

Our story illustration framework uses tags – image manually inserted textual annotations – to compare images with news articles. Other automatic methods could also provide us with the textual description of images (colors, shapes, textures). Image textual descriptions (or annotations) are necessary to compute the degree of association between images and news sentences. However, even with a well annotated image database available, choosing images which best represent a sentence



is not a simple task. The problem is subjective, and a human would have to make an empirical assessment, using knowledge gained over the years, to judge the significance of each image. Our framework follows this type of approach. However, we propose a finer grain approach aiming at illustrating individual sentences and not full stories.

## 3.2 Standard Processing

To implement the ideas proposed in this work and effectively compare the most important news terms with image annotations, an Informational Retrieval (IR) model was used. An IR model consists of a document model, a query model and a model for computing similarity between documents and queries. One of the most popular IR models is the vector space model and various effective retrieval techniques have been developed according to this model (Salton and McGill 1986; Buckley and Salton 1995; W. M. Shaw 1995). Among them, term weighting and relevance feedback are of fundamental importance. This chapter focuses on the first while Chapter 4 focuses on the second.

In our framework, news sentences are used to search for relevant documents and the database from where documents are retrieved is represented by the collection of images. Both news sentences and images are represented as vectors. Our proposed model computes the degree of similarity between vectors and obtains a rank for each image candidate for illustration. This process is divided in three steps.

### Text Processing

In the first step of the illustration process, news articles are segmented into sentences and stop-words – words that carry little information (“the”, “and”, “is”, etc) – are removed using a list of English language stop-words, retrieved from the Apache *Lucene* search engine, as reference. *Lucene* is an open source information retrieval software library and was only used as the source of a stop-word list. At this point, sentences will be the unit of text-to-image comparison even though the actual prototype groups sentences with less than 80 words with the next sentence. With this method, short sentences are not used in the image selection algorithm and more terms are used in this process. After this procedure, sentences are represented by their terms.

In linguistic morphology, stemming is a process for reducing words to their base, stem or root form. The new word need not be identical to its morphological root, as it is usually more accurate to relate words to the same stem, even if the stem is not itself a valid root. For example while words similar to “cats” and “catlike” are reduced to the root word “cat”, while words like “demonstrable” and “demonstrate” are stemmed to “demonstr”. This has been a problem in computer science for a long time, and the first paper on the subject was written by Lovins (1968). In 1980, Martin (Porter) published another stemmer, one that became the standard algorithm used for English language. The application described in this work uses this process in both image tags and news terms before the next step because using stemmed word improves the similarity calculations.

### **TF-IDF Weighting**

In every sentence, some words define the content and are therefore, more important. For the purpose of ranking words by their importance, the weighting technique term frequency-inverse document frequency (tf-idf) is applied. The tf-idf weight is a statistical measure used to evaluate how important a word is to a document in a collection. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the collection of documents. In the second step, both terms and tags are weighted according to their frequency in the respective news article or image annotation set and to their frequency in the collection of news or images. Usually, only text document are used, but it was adapted to our case, meaning that a document is a set of news terms or image tags.

If we define  $w_{i,k}$  as the weight for term  $t_k$ ,  $k = 1 \dots N$  in document  $i$ , where  $N$  is the number of terms, document  $i$  can be represented as a weight vector in the term space:

$$D_i = (w_{i,1}, \dots, w_{i,k}, \dots, w_{i,N})$$

To correctly estimate the weights, we need to consider two aspects: First, if term  $k$  is frequently occurred in the document  $i$ , then  $w_{ik}$  should be assigned high value. This intuition suggests that a term frequency ( $tf$ ) factor should be included in the estimation of  $w_{ik}$ ; second, tf alone cannot ensure an acceptable estimation. When

the high frequency term is not concentrated in a few documents, but instead spreading over all documents, we should give this term low weight. This introduces the inverse document frequency (*idf*) which varies inversely with the number of documents in which a term appears. The *idf* formula is

$$idf_k = \log\left(\frac{|D|}{df_k}\right),$$

where  $|D|$  is the total number of documents in the collection and  $df_k$  is the document frequency for term  $k$ . If the term is not in the corpus,  $df_k = 0$  could lead to a division by 0. It is therefore common to use  $df_k + 1$  but in our case there is no need because only terms in the collection of documents are used. Experiments have shown that the product of *tf* and *idf* is a good estimation of a terms importance (Salton and McGill 1986; Buckley and Salton 1995; W. M. Shaw 1995). The final formula to weight each term is

$$(tf-idf)_{i,k} = tf_{i,k} \times idf_k = \frac{n_{i,k}}{\sum_j n_{j,k}} \times \log\left(\frac{|D|}{|\{d: t_k \in d\}|}\right),$$

where  $n_{i,k}$  is the number of occurrences of the word  $k$  in the set associated with the document  $i$  (usually  $n_{i,k} = 1$  for image tags),  $\sum_j n_{j,k}$  is the number of words in the same set,  $|D|$  is number of documents in the collection and  $|\{d: t_k \in d\}|$  represents the number of documents with the term  $t_k$ .

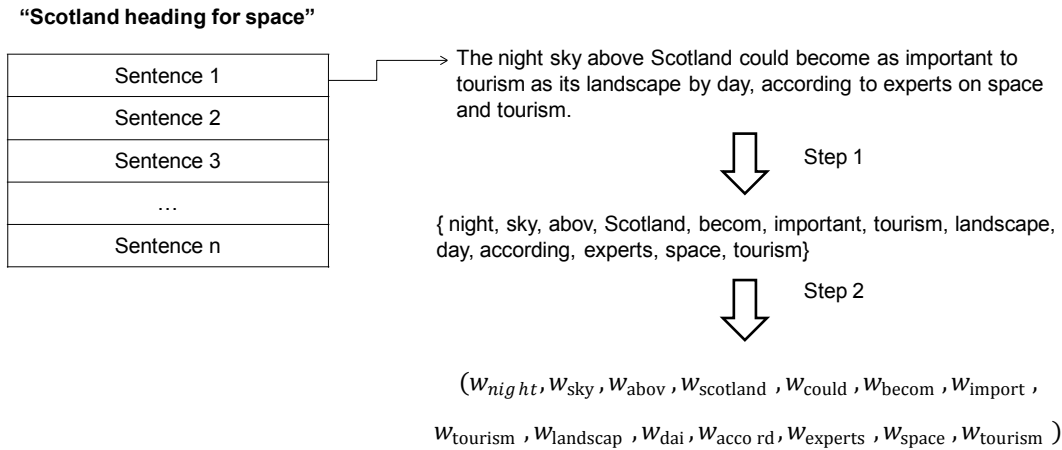
After this step and reminding that in our framework news are represented by sentences, a sentence  $n$  from news  $m$  is represented by the vector

$$s_{mn} = (w_{mn,1}, \dots, w_{mn,T}),$$

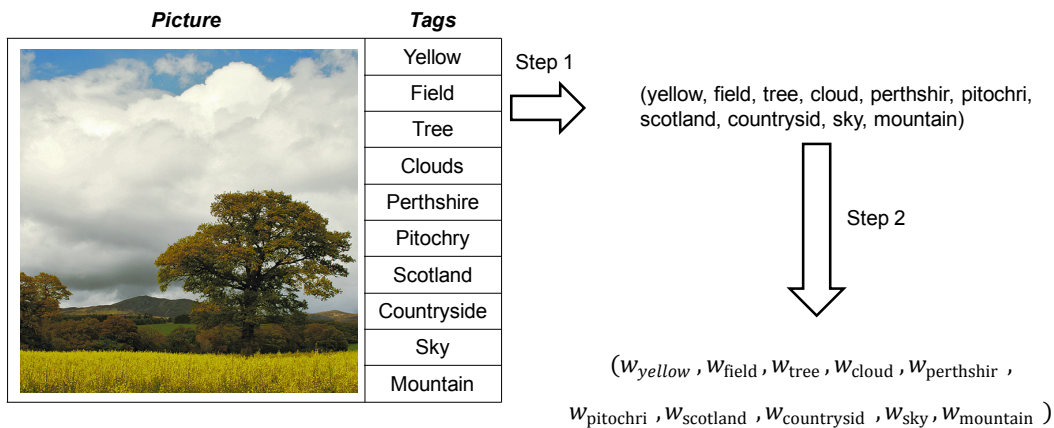
where each component  $w_p$  indicates the *tf-idf* weight of the respective term, from a total of  $T$  news text terms. An image is represented by the vector

$$i_k = (w_{k,1}, \dots, w_{k,T}),$$

where component  $w_p$  indicates the *tf-idf* weight of the respective term, from a total of  $T$  image tags. Both sentence and images are represented by vectors in the same dimensional space.



**Figure 3.1: Text processing (Step 1) and tf-idf weighting (Step2) applied on a sentence.**



**Figure 3.2: Text processing and tf-idf weighting applied on a picture.**

### Sentence-Image Comparison

In the third step the degree of association between an image and a sentence is computed, taking into account weighted image tags and sentence terms. Note that in this step the final product is an image that best matches a particular sentence. Thus, images need to be ranked according to their relation to the sentence at hand. An image similarity rank is computed with the cosine similarity (or cosine distance) which measures the angle between two vectors. The similarity between vectors of  $n$  dimensions is found by the cosine of the angle between them. It is a technique often used in information retrieval, more specifically in text mining and is given by the equation

$$D_{\cosine}(s_{m,n}, i_k) = \frac{s_{m,n} \times i_k}{\|s_{m,n}\| \times \|i_k\|}$$

where  $s_{m,n}$  is the  $n^{th}$  sentence of the  $m^{th}$  news article, and  $i_k$  is  $k^{th}$  image. With this technique, it is possible to estimate the degree of relation between an image and a sentence.

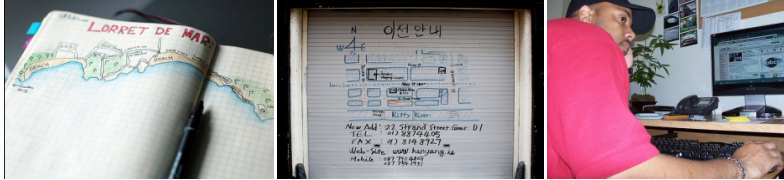
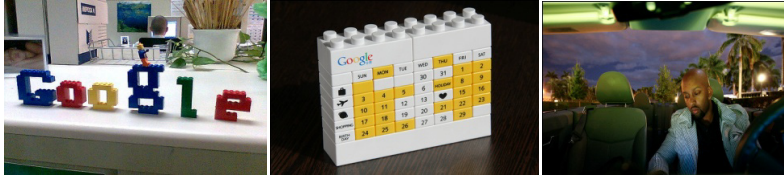

Sentence	Retrieved images
Internet mapping is wiping the rich geography and history of Britain off the map, the president of the British Cartographic Society has said.	
Mary Spence said internet maps such as Google and Multimaps were good for driving but left out crucial data people need to understand a landscape.	
Mrs. Spence was speaking at the Institute of British Geographers conference in London.	

Table 3.1: Illustration of “Online maps ‘wiping out history’”.

Table 3.1 shows an example of three images with highest rank and therefore, primary candidates to illustrate the text. In this example, it is possible to see cases of good and bad selection. In the first two rows the relation between image and text is clear: – in the first, “maps” and “cartography” relate to the first two pictures and “internet” to the third one; in the second sentence, “Google” stands out in two pictures and the other has a man driving a car. The table’s last row has an example where the pictures may not seem adequate to illustrate the current text. But when viewing the pictures annotations we see that “London” is part of the set and depending on the number of pictures tags (if they are few), the keyword London may have a high classification.


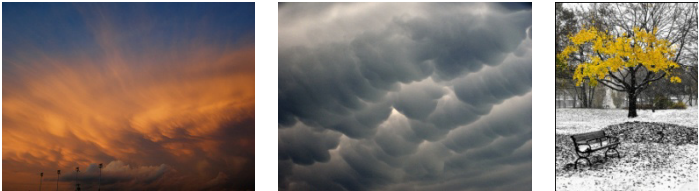

Sentence	Retrieved images
<p>Spread by infected midges, the disease is hard to control and causes a painful death in the sheep that are affected.</p>	
<p>What makes British farmers different to the French, Belgian and Dutch? That answer is a combination of luck, the weather and science.</p>	
<p>But this week the winds have begun to shift to the south with the potential to sweep in bluetongue-infected midges from the continent.</p>	

Table 3.2: Illustrations of “UK escapes the worst case of bluetongue”.

In Table 3.2 good examples of sentence illustration are presented, but surprisingly the best ranked images only have one word in common with the text. It is a recurring situation in which images are selected because of a news term with a high weight. One problem with this process is the fact that only images that have the exact text words can be selected for illustration. A solution to this problem is presented in the next section. This brings forth a limitation in our framework as only exact matches (tags and terms with at least the same root word) are associated. A solution to this limitation is proposed next.

### 3.3 Semantic-based Comparison

In the past, when we did not know the definition of a certain word we would look in the dictionary. Nowadays, we “Google-it” to find the words meaning in online dictionaries. In addition to the words definition, these online services, as in traditional dictionaries, also provide a list of synonyms of the word. The defined framework for comparing news text to image tags is limited to their terms and tags, leaving out the true semantic interpretation of both elements. It is possible to use

dictionary tools when searching for illustrations that best fit the text keywords. Thus, the search for a term can be widened to its synonyms. In our framework we tackle this challenge by expanding sentences and image tags with a linguistic ontology (WordNet<sup>4</sup>) to enclose all possible meanings. WordNet is a large English database that groups nouns, verbs, adjectives and adverbs in sets of cognitive synonyms, interlinked by means of conceptual-semantic and lexical relations. Another characteristic of WordNet is the use of hierarchical organization to induce a transitive relation among words. Considering the word *democrat*, we have *democrat* → *politician* → *leader* as its hierarchy, with *democrat* as the lowest and *leader* as the highest word in the hierarchy. Using WordNet we can relate text terms with their hypernyms – words whose semantic field includes the word being used (e.g. *animal* is hypernym of *dog*) – and do the same with image annotations in order to compare words at a higher hierarchical level.

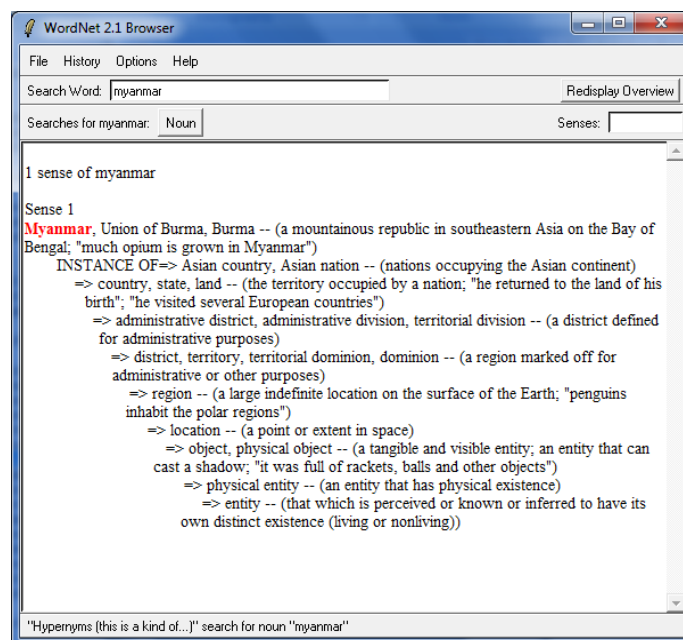


Figure 3.3: WordNet browser.

WordNet was used to create a group of synonyms and hypernyms. This process is applied only to nouns. For every noun  $t_i$  in the sentence vector  $s_{m,n}$ , we use WordNet to compute its synonyms,

$$Syn(t_i) = (ts_{i,1}, \dots, ts_{i,T}),$$

<sup>4</sup> <http://wordnet.princeton.edu/>

where each  $ts_{i,k}$  corresponds to a synonym of term  $t_i$ . Synonyms will have the same weight as the original term and there are no repetitions in the resulting vector. The original term is also included. Thus, the resulting sentence is

$$s_{m,n} = (t_1, \dots, t_T) + \sum_{i \in nouns(t_1, \dots, t_T)} (ts_{i,1}, \dots, ts_{i,T}),$$

where the first part of the equation represents the original sentence and the second represents the sentence-nouns synonyms.

A similar result is obtained using the hypernyms function, where one word becomes a vector

$$Hyp(t_i) = (th_{i,1}, \dots, th_{i,T}),$$

where each component  $th_{i,k}$  corresponds to a hypernym of  $t_i$ . Again, hypernyms will have the same weight as the original term, which is also included in the sentence vector. When expanding hypernyms, the root of the concept hierarchy of a noun corresponds to the word “entity” or “abstract”. This would produce similar results for very different sentences. We propose a resolution to this predicament: If a noun is  $N$  hypernyms from the root we only consider the first  $N/2$  hierarchy levels; for example, the word “continent” hierarchical tree: continent → landmass → land → object → physical entity → entity – would be transformed in (continent, landmass, land).

The resulting sentence after using the hypernym function is

$$s_{m,n} = (t_1, \dots, t_T) + \sum_{i \in nouns(t_1, \dots, t_T)} (th_{i,1}, \dots, th_{i,T}),$$

where the first part of the equation represent the original sentence and the second represent the sentence-nouns hypernyms.

A third option to perform the semantic expansion is to consider both synonyms and hypernyms. Thus, a sentence is then represented by its terms and the synonyms and hypernyms of its nouns

$$s_{m,n} = (t_1, \dots, t_T) + \sum_{i \in nouns(t_1, \dots, t_T)} [(ts_{i,1}, \dots, ts_{i,T}) + (th_{i,1}, \dots, th_{i,T})],$$



where the first part of the equation represent the original sentence and the second represent the sentence-nouns synonyms and hypernyms.




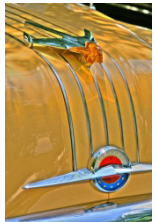







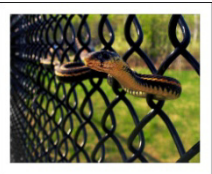
Sentence	Tf-Idf	Synonyms	Hypernyms	Synonyms + Hypernyms
<p>“The Yellow School Bus Commission, chaired by former Education Secretary David Blunkett, said it could "revolutionize" the school run.”</p>				
	Yellow	Yellow; Autobus	Car; yellow	Automobile; Yellow
<p>“The MSc Ethical Hacking and Computer Security course at Abertay University will explore the methods criminals use to attack networks.”</p>				
	Computer	Course; Track	Computer	Course; Track
<p>“A three-meter (10-foot) python has killed a student zookeeper who let the snake out of its enclosure in Venezuela while working a night shift at the zoo.”</p>				
	Snake	Snake	Vertebrate; Reptile; Zoo	Vertebrate; Reptile; Snake; Zoo

Table 3.3: Illustrations generated by the different illustration methods.

Table 3.3 shows results provided by the three new methods of association when compared with the original tf-idf based method. It shows instances where the new methods increase the quality of the returned images and others where the resulting images are the same. The method that uses a semantic expansion of both synonyms and hypernyms in some cases returns the same results as the synonyms or hypernyms methods, depending on the method with higher ranking results. Recall

that when computing the similarity between images and sentences, the number of tags similar to the terms is as important as having matching words with higher ranks.

### 3.4 Story Sequence Consistency (SSC)

Because reading is a sequential process, one needs to keep a sequential consistency between illustrations to guarantee a semantically coherent set of images. To guarantee that the selected illustrations have a similarity not only with one sentence, but with the content of the news article as well, we employ a memory based function to consider the previous sentences. A sliding window parses neighboring sentences to calculate the accumulated weight of each news term. The final ranking function

$$Rank(s_{m,n}, i_k) = \sum_{p=n-Windowsize}^n \left[ \frac{1}{n-p+1} D_{Cosine}(s_{m,p}, i_k) \right],$$

computes the rank position of image  $k$  in relation with the  $n^{th}$  sentence of the  $m^{th}$  news article. The variable  $Windowsize$  indicates the windows range to include the previous sentences. The factor  $(n-p+1)^{-1}$  is a weight decay to adjust the contribution of sentences according to their distance to the  $n^{th}$  sentence.

Figure 3.4 illustrates the purpose of the sliding window algorithm: the image-sentence association algorithm takes into account not only the sentence currently being displayed, but also the previous sentences while giving more importance to the current sentence.

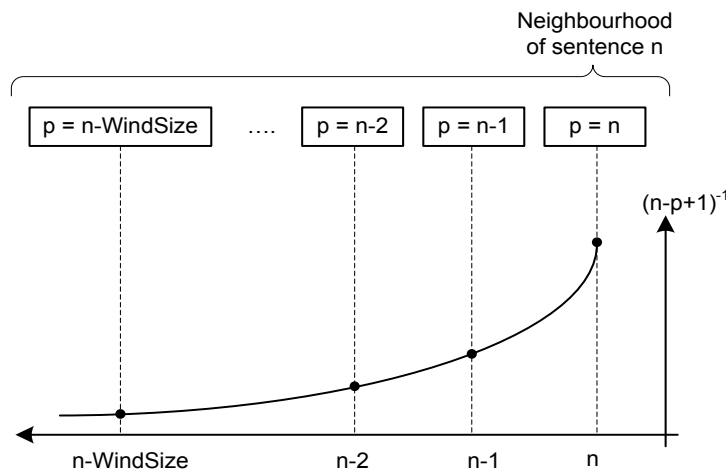


Figure 3.4: Sliding Window on Story Illustration application.

## 3.5 User Evaluation

To evaluate our approach and to assess the different illustration methods we conducted a user evaluation. For this purpose we used a dataset of BBC<sup>5</sup> news downloaded from their website and a dataset composed by Flickr<sup>6</sup> images.

### 3.5.1 News Dataset: BBC Web News

News articles were collected from the BBC website. A total of 6.727 news articles were collected with each belonging to just one subject and section. On the BBC website news URLs are organized according to these two categories. Therefore the news subject and section is available for extraction from the article's URL. Each news category is obtained via assignment by BBC journalists.

#### Data Pre-processing

Since the files from BBC are in *.xhtml* format, a parser was used to extract the needed information. From the many existing xml parsers, two of them are the most generally known: DOM (Document Object Model) and SAX (Simple API for XML).

DOM is a tree-traversal API (Application Programming Interface), best suited for applications that need to access the document repeatedly and out of order. For this end, it loads the entire document which would make our application run slower unnecessarily. In our case only one survey is necessary for each document.

SAX was more adequate API because it is a serial access parser that uses callback methods called whenever an event occurs, for example, when an element begins or ends. The problem is that it was designed for XML and while it is similar to XHTML, it is not exactly the same. Special characters like bullets are represented differently in each language.

A SAX parser was developed to remove navigational content and extract the news corpus and title. Special care was taken to handle language specific characters and other formatting data. Steps 1 and 2 of the standard processing, described on section 3.2, were applied to the retrieved text. The resulting data was stored in a database.

---

<sup>5</sup><http://www.bbc.co.uk/>

<sup>6</sup><http://www.flickr.com/>

### 3.5.2 Images Dataset: Flickr Photos

A collection comprising 25.000 Flickr images was redistributed by Huiskes and Lew (2008) for research purposes. They extracted the image tags and other image metadata. The tags were inserted by Flickr users with a folksonomy which in some cases, results in insufficiently described images. Alternatively, some tags are not always written correctly as they were not inserted by professionals. From the whole dataset, 20.000 had at least five tags; thus, the remaining images were not used in our experiments.

#### Data Pre-processing

An application was developed to store image data in a database. The application parses the annotation files and saves both image tag and path in the database. It also associates each tag with respective weight, calculated as described in section 3.2. At this point, steps 1 and 2 of the standard process were already performed on the image annotation sets.

### 3.5.3 Data Model

News data and pictures were saved in a SQLite<sup>7</sup> database. This database is particularly adequate for this application as it implements a standalone, transactional SQL database, with zero-configuration and platform independent. Application data such as section, headline, news text and tags are saved in the database.

Tables Table 3.4, Table 3.5 and Table 3.6 illustrate examples in the tables stored in the SQLite database: the first table contains news subject and section, headline and news text; the second table contains image tags processed with the stemming algorithm (stemmed tag), the associated pictures location (image path) and the respective weight; the third table contains the news terms and their weight associated with the stemmed word and the respective news title.

---

<sup>7</sup> <http://www.sqlite.org/>

ID	Subject	Section	Headline	Text
1432	Sports	Scotland	TV rights boost for Scottish FA	“The Scottish football association has agreed a new television deal...”
2234	Education	Bradford	Ex-burglar ‘offered’ study place	“A Bradford student rejected by a London medical college...”
323	Business	Scotland	‘No road grit left’ says council	“The Scottish government has insisted there are "very substantial"...”
4423	World	Americas	Lula urges Brazil Olympic boost	“Brazilian president Luiz Inacio Lula da Silva has said his country needs...”
532	World	Africa	Zimbabwe call parties call for peace	“Zimbabwe’s ruling and opposition parties have...”
3232	World	Americas	US airline ‘broke safety rules’	“US aviation officials are accusing American Airlines of...”

Table 3.4: News articles table.

Tag ID	Tag	Stemmed Tag ID	Stemmed Tag	Rank ID	Rank	Image path
3423	Hip	3142	hip	3128	0.6947	../db/Flickr/im7731.jpg
8984	Smoke	8543	smoke	12654	0.4432	../db/Flickr/im19378.jpg
7563	Stick	7476	stick	12795	0.2975	../db/Flickr/im19348.jpg
12464	Tattoo	11463	tattoo	3129	0.4908	../db/Flickr/im7731.jpg
12464	Tattoo	11463	tattoo	3130	0.8998	../db/Flickr/im9382.jpg
31603	White	31550	whit	4644	0.7786	../db/Flickr/im19378.jpg

Table 3.5: Pictures table.

Term ID	Term	Stemmed Term ID	Stemmed Term	Rank ID	Rank	Headline
423	Azerbaijan	411	azerbaijan	655	0.0409	Georgia falls victim to pipeline politics
923	Challenge	901	challeng	22	0.0405	Public finances improve slightly
10342	Other	10032	other	32	0.0037	Back pain eased by good posture
32324	Wait	30320	wai	143	0.0160	Shunned Starbucks in Aussie exit
44233	Zebra	42713	Zebra	5665	0.0702	Animal extinction on central Africa

Table 3.6: News words table.

In the news table, news articles are associated with a subject and a section. In the pictures table, both keyword and picture location can be repeated, since one picture tag can be associated to multiple pictures and one picture can have multiple tags associated, as we see on various websites like Flickr.

### 3.5.4 Experiment Design

To assess the illustration methods we conducted a user evaluation with twenty-three subjects (five females and seventeen males). The subjects were computer science academics and graduate students. Table 3.7 presents the age demographics of the group of users. Again, the user study was performed on a Windows machine and all users had headphones to isolate ambient noise. The usage of the Text-to-speech feature was optional as it was not under evaluation. The objective of the evaluation was to determine the most accurate illustration method.

The methods under evaluation were:

- **Synonyms** – Synonyms of the text nouns are used to expand the queries;
- **Hypernyms** – Hypernyms of the text nouns are used as a query expansion;
- **Synonyms + Hypernyms** – Synonyms and hypernyms of the text nouns are used as a query expansion;
- **SSC** – Previous sentences of the news article are used as a query expansion;

Age	
Min	22
Max	43
Median	27.22
Mean	25

Table 3.7: Test users' age.

The tests were split into two parts:

- **Step 1** – Subjects were asked to give general information (age, gender and background);
- **Step 2** – Subjects read six news articles, each with a combination of a specific illustration method with the story sequence consistency feature on or off. After watching each multimedia presentation the tester had to rate the method;

Users watched news audio/visual presentations while rating the illustration methods. In the end, they were asked to give comments and suggestions regarding the application.

### 3.5.5 Results and Discussion

Subjects rated between 1 (worst) and 5 (best) to each method after watching one article with a particular method.

The SSC method presented better results than using a single sentence – SSC with hypernyms was the best method. With synonyms, the semantic expansion is limited to the same objects; consequently the sentence-image relation is not greatly improved. Alternatively, using hypernyms actually widens the semantic meaning of a sentence, capturing richer sentence-image relationships.

Illustration method	Normal	SSC
Synonyms	2,45	2,88
Hypernyms	2,38	2,94
Synonyms + Hypernyms	2,56	2,25
Average	2,46	2,72

Table 3.8: User assessment of the illustration methods.

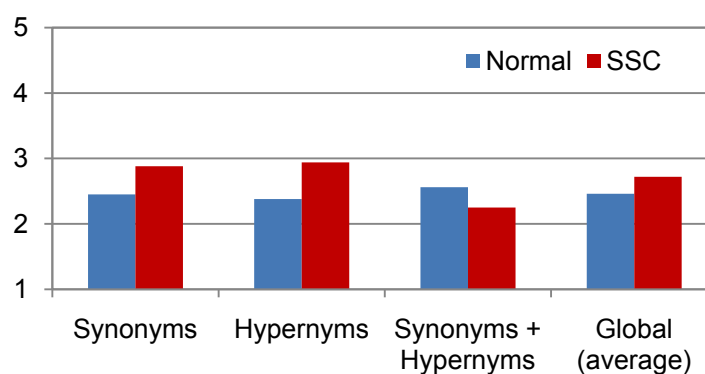


Figure 3.5: User assessment of the illustration methods.

In some cases, a few of the selected illustrations might be unrelated to the news content and the connection with the text can be hard to understand. This happens in some situations due to external factors as improper image tagging. Using both synonyms and hypernyms should have improved the quality of the selected images, but instead, as we can observe in Table 3.8, results show the opposite. This is related to a too wide semantic expansion of the initial nouns. Other lessons were learned from the experiment. Bad image tagging creates discrepancies in the weighting method. Words that appear in few sets and of small length (sets with few annotations) tend to have higher rank. For example, a picture shows a dog jumping towards another; it only has the tag “attack” associated. If we calculate the resulting weight we see that the result is

$$\left(\frac{\text{number of tag occurrences in set}}{\text{total number of tags in set}}\right) * \log\left(\frac{\text{total number of images}}{\text{number of images with the tag}}\right) =$$

$$= \left(\frac{1}{1}\right) * \log\left(\frac{25000}{1}\right) \approx 10,127,$$

which is an unrealistic value; Of course this particular picture would never be selected due to our minimum tags restriction; but there are cases where pictures with few tags, have high ranking words associated, and with slim relation to the picture, which makes the image a candidate for illustrating news of completely different context.

Some sentences aren’t complete. The SAX parser still needs perfecting because it can’t read all of the news files since they don’t all follow a standard template. For example, sport score news only show game table results and therefore are harder to parse than news of other sections that are usually just plain text. For this reason, not all news files were used in the experiment and some of the sentences retrieved were stored in the same format as they were presented in the online version.

The list of English stop-words used also needs further improvement; sometimes in the News reading debug mode we can see words of low importance being used.



### 3.6 Summary

This chapter described the proposed process of automated sentence illustration. The Story Illustrator standard procedure performs basic text processing on sentences and image tags and a weighting technique classifies words by their importance. The technique to compute the similarity between illustrations and sentence was carefully described. Also, an ontology was explored to refine the sentence-image relationship and a method to improve the coherence between sequentially selected illustrations was implemented. The user tests evaluated the introduced illustration methods which result from a combination of the ontology comparison with the story sequence consistency, on top of the standard processing. Results show that hypernyms with SSC were the best sentence-image association method and general user opinion confirmed the hypothesis that illustrations are both useful and enjoyable. Illustration methods with the SSC algorithm proved to be the users' preferred method for selecting illustrations for a given sequence of sentences.

Besides the results, experiments also returned important information on how to improve the framework and how to design improved user tests in this novel domain.

Following this chapter, a paper was published presenting the described work. User experiments and respective results were also reviewed on the paper (available in the annexes):

- Diogo Delgado, João Magalhães, Nuno Correia, “Automated illustration of news stories”, IEEE International Conference on Semantic Computing, Pittsburgh, United States of America, September 2010.



## Chapter 4

# User Feedback based Illustration

Information retrieval systems are used to reduce what has been called as “the information overload”. Too much information can confuse users and hinder their decision process. As the world progresses to a ubiquitous information age an increasing number of people are connecting to the Internet. Users are an integral part of the Internet through their implicit feedback when clicking a link, submitting a review, or tagging an image. This interaction is helpful to isolate relevant information by analyzing the documents they navigated, i.e., their implicit feedback.

In this chapter we will study how feedback can be used to personalize the selection of the sentence illustrations for news stories.

### 4.1 Related Work

Relevance feedback (RF) techniques used in IR are popular techniques to refine search results. RF is a formula where the user can guide the retrieval process by interactively updating the search query. This interactive approach differs from fully-automatic approaches where retrieval is performed by fixed weighted combination of features, and includes the user in the loop of the retrieval process by dynamically and interactively updating the weights applied to different feature vectors. Two types of RF are exploited in this work: explicit and implicit feedback. Explicit RF generally separates the relevance of a particular image to a query in two categories: positive (relevant) and negative (not-relevant). Some systems adapt relevance feedback with a performance comparison by adding multi-class methods

(Peng 2003). Various RF systems estimate the ideal query parameters based on only low-level image features such as color, texture, and shape. With a few positive and negative example images, the RF system is able to improve reasonably the results' accuracy. On the other hand, if the user is searching for a specific object that cannot be sufficiently represented by combinations of available examples, these systems will not return relevant results. To address the limitations of these feedback systems, Lu, Hu et al. (2000) proposed a framework that performs relevance feedback on both low-level feature vectors and images annotations.

Implicit feedback techniques includes any kind of natural interaction of a user with a document (White, Jose et al. 2006). Examples of implicit feedback are mouse and keyboard actions, page navigation, bookmarking, and display time. There is still some debate over the value of implicit feedback but numerous controlled studies suggest that there is a clear correlation between the time a user spends viewing a document and how useful they found that document (Goecks and Shavlik 2000; Claypool, Le et al. 2001; Fox, Karnawat et al. 2005). Scrolling behavior is simple to observe and it is easy to compute the display times of different document segments. Accurate feedback from display time would be highly valuable since it usually logged by browsers. In our work, it is not possible to extract implicit feedback through scrolling behavior because of the sentence-by-sentence fashion presentation. Instead, we propose a different approach to infer implicit behavior: using user's reading speed we can infer their attention levels. We calculate the reading speeds by counting the average number of words read per minute.

Information Retrieval new major research topics revolve around a more complex environment with user personalized adaptive information retrieval or filtering and recommendation systems. As opposed to traditional ad-hoc systems, personalized systems profile each user, automatically adjusting to the user's specific preferences. To profile a user, a personalized system usually needs a significant amount of explicit feedback (training data) from the user. However, the average user prefers not to answer a lot of questions. Simultaneously, a user does not want to endure the initial poor performance provided by the system in "training" and expects a system to work reasonably well as soon as the first uses of the system. Good initial performance is an incentive for the user to continue using the system. Thus an important aspect of personalization is to develop a system that works well initially

with less explicit user feedback. Zigoris and Zhang (2006) use a combination of both explicit and implicit feedback to build a user's profile and uses Bayesian hierarchical methods to borrow information from existing users. In our framework we do not make use of user personalization, leaving this as future work.

Several approaches exist to infer user feedback from their interactions, however, very few works have explored the affective analysis of user behaviors as Lee, Chang et al. (2007). They presented HiTV, an emotionally-reactive TV system that reacts to social responses. The viewer has a soft ball that serves as an interface, from which it is possible to interact with the television through emotional reactions like holding tight or shaking it. The system interprets these actions and responds with visual or audio amplifications.

## **4.2 Personalized Sentence Illustrations**

In our system, users can be unsatisfied with the illustrations even when pictures are accurately associated with sentences. It is a choice that depends on one's preferences. With this goal in mind, two new relevance feedback techniques were added to the framework: explicit feedback and timed feedback.

In traditional relevance feedback, users provide new queries with results. The system uses the modified query in the ranking functions to compute the next set of results. In our system, the approach is slightly different because users do not provide new examples or queries; instead, users rate the selection that the system made for them. Thus, the system needs to personalize the illustrations according to the user ratings of the selected illustrations and the time they spend reading each sentence.

### **4.2.1 Explicit Feedback**

Relevance feedback is the process of adjusting an existing query using information provided by the user about the relevance of previously retrieved documents. In the image retrieval field, classifying pictures as good or bad provides the system with an additional help in the selection process. To implement the explicit relevance feedback and since we use a vector space model, an adaptation of the Rocchio feedback algorithm was used. In this approach the original query is revised to include a fraction of relevant and non-relevant documents. In our framework, this is obtained by increasing the weight of relevant tags and decreasing the weight of

non-relevant tags. If the set of relevant image tags ( $D_P$ ) and non-relevant image tags ( $D_N$ ) are known, it is possible to reach an optimal query with the following equation

$$Q' = \alpha Q + \beta \left[ \frac{1}{N_P} \sum_{i \in D_P} D_i \right] - \gamma \left[ \frac{1}{N_N} \sum_{j \in D_N} D_j \right]$$

where  $\alpha$ ,  $\beta$  and  $\gamma$  are suitable constants;  $N_P$  and  $N_N$  are the number of images in sets  $D_P$  and  $D_N$  and  $Q$  is the set of sentence terms. Recall that the system uses sentences to find relevant illustrations. As the iteration progresses, better results are returned.



**Figure 4.1: News Reading mode (highlight in the feedback functionalities).**

In the news reading mode (Figure 4.1), readers can rate a picture as good or bad with the buttons *Like* and *Don't Like*, depending if it is adequate to the content of the story and if readers simply like the picture. When a picture is marked (as good or bad), the associated tags are stored in a list (of positive or negative) tags. The changes take effect in the next iteration which means the next sentence will be modified. For example, an image with the tag *sea* illustrates a sentence with the same term; if the reader marks this picture as good, the set of image keyword tags will be added to a list of positive images. The next query (sentence) will be modified according to the Rocchio formula and images with *sea* as an annotation will have their rank reinforced. Alternately, if the picture is marked as bad, the annotations will

be added to a list of negative images. These two options should be used when the picture has at least a slim relation with the content of the text.

In some cases the reader may feel there is no relation between pictures and text. For these cases, an option was added: the *Inadequate* button. When used the procedure is the same as the *Don't Like* button (tags are stored in a list of negative tags) but also the sentence terms are added to the pictures' set of tags with a negative weight. This will cause the picture not to be selected again for illustrate sentences with the same terms. As an example, a user is reading a news article portraying Israel attacks on the Gaza strip and an image of a dog attacking another appears; the user may not understand that the word "attack" had a large portion of the similarity calculations but still it is not an adequate picture for this article; choosing the inadequate option will lower the importance of the image tags and add the sentence keywords to the picture. In the next viewings of this news article (or others with similar contents) this picture will have very low probability of being selected for illustration.

#### **4.2.2 Timed Feedback**

To assess users' reading speed and still modify the query with information provided by user feedback, a method for collecting feedback was proposed: the timed feedback. This method is an adaptation of the relevance feedback with a timed element. The application counts the time a user takes to read a sentence and associates the number of words read per minute with an average. In the following iterations, the reading speed is calculated again and the distance to the average speed is measured. A window is computed with the average as its center and if the distance between current and average reading speed is within the window size, the application will proceed as the relevance feedback method. Otherwise when the picture is rated as relevant or non-relevant only a fraction of the terms' weight will be accounted. For example, using a window of size 200 words per minute (wpm):

- In the first sentence, the user reads the sentence at 200 wpm and rates the illustration. The application will proceed as the relevance feedback – the sentence terms are stored in a list of positive terms;
- In the second sentence, the user reads at a speed of 150 wpm and again presses a user feedback button, marking the sentence's illustration. The

reading time is between 100 wpm (average reading time minus half the window size) and 300 wpm (average reading time plus half the window size) therefore, the applications will do the same procedure as the relevance feedback method. The average reading speed is updated;

- In the third sentence the user rates the picture again but at a reading rate of 600 wpm. This proves the lack of attention that we are trying to filter. Consequently only a fraction of each terms' weight is added to the list of positive or negative tags;

With this method, we pretend to separate user emotional disposition and only trust judgments in cases of user attention. The question that arises is how to judge the first iteration: If a user reads the first sentence and marks the association with a too high or too low reading rate, how to proceed with this information? We tackle this limitation with a solution based on an assumption that the average person's reading speed is from 120 to 200 wpm – values obtained by general research. In the first iteration, the user's reading speed is compared with the proposed window and the algorithm advances as before. Figure 4.2 shows an example of a user reading speed – words per second were used to simplify the example. In our proposed feedback method, only values within the windows range will be used with the Rocchio technique therefore in the example, the rating of the illustrations from sentence 4 and 5 will be less relevant in next iterations.

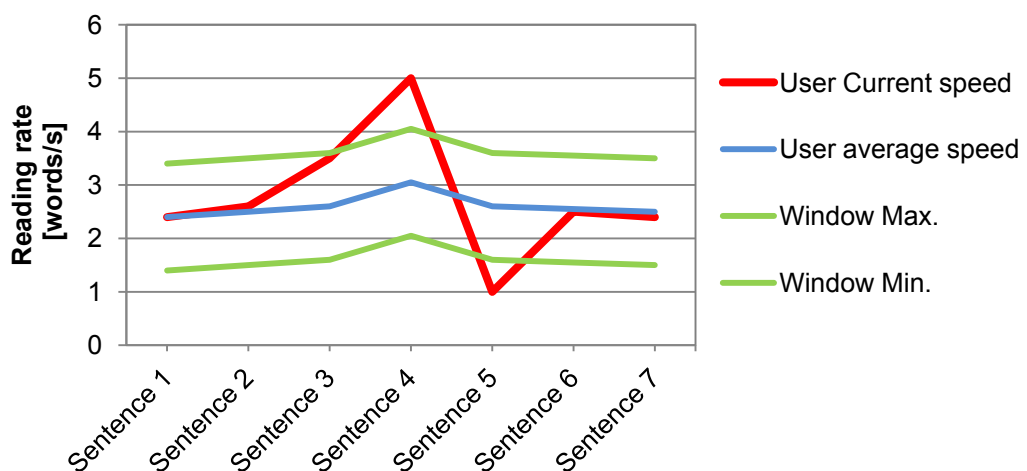


Figure 4.2: Relevance feedback example of a user reading speed.



### 4.3 Improved Story Sequence Consistency

After the experiments presented in the previous chapter, we concluded from user comments that some of the automatic sentence illustration methods could be modified for improving the illustrations' adequateness to the full story. We decided to change the calculations for the selection of images by integrating the title and the full text. Now, for any given sentence, the titles and all sentences are taken into account, with a lesser weight than the result produced by the previous ranking function. Having  $Rank(s_{m,n}, i_k)$  as the previous image rank, we compute the title's rank as

$$Rank(s_{m,headline}, i_k) = D_{Cosine}(s_{m,headline}, i_k),$$

while the rank of the full news article is calculated as in a sentence by sentence order

$$Rank(s_m, i_k) = \sum_{j=1}^n D_{Cosine}(s_{m,j}, i_k),$$

which results in the final ranking function

$$RankFinal(s_{m,n}, i_k) = \delta Rank(s_{m,n}, i_k) + \varepsilon Rank(s_{m,headline}, i_k) + \theta Rank(s_m, i_k).$$

In the final ranking equation,  $\delta$ ,  $\varepsilon$  and  $\theta$  are suitable constants with the respective values of 0.65, 0.15 and 0.20. The previous rank value represents 65% of the final ranking function, the headline rank 15% and the sum of all sentences 20%. These values were found empirically. They attempt to give relevance to the title and news content in the similarity calculations but in a smaller scale.

### 4.4 User Evaluation

To evaluate the personalization of sentence illustrations with the two user feedback techniques we conducted a user evaluation to assess their success. The data used in this evaluation is the same as the one we used in the previous chapter. The test subjects were different from the previous evaluation.

#### 4.4.1 Experiment Design

To assess the specific feedback methods we conducted a user evaluation with nine subjects (one female and eight males). The subjects were academics and graduate university students. Table 4.1 presents the age demographics of the group of users. Again, the user study was performed on a Windows machine and all users had headphones to isolate ambient noise. The TTS feature was optional in these tests as the feature was not under evaluation. Users read news articles using the illustration method *hypernyms* with the improved SSC. The illustration method was selected based on the scores and comments obtained on the chapter 3 tests.

Age	
Min	22
Max	25
Median	23,89
Mean	24

Table 4.1: Test users' age.

The test order of the feedback methods was also carefully distributed across tests to remove the bias toward the first tested methods, which is when users are less familiar. In each user test, users read two news articles using each of the three methods under evaluation. Also, users read one of these two articles with the three methods (Table 4.2). A total of six articles were read in each test and the evaluated methods were:

- **No Feedback** – User-given image rates are stored but are not taken into account when associating images to sentences;
- **Explicit Feedback** – In each iteration, user-given image rates are stored and are taken into consideration when searching for next illustrations;
- **Timed Feedback** – The application stores user-given image rates and the time the user takes to mark the images. A reading speed is calculated and, in combination with the image rate, is taken into consideration when searching for next illustrations;

	News 1	News 2	News 3	News 4	News 5	News 6
User 1	123	1	2	3	0	0
User 2	0	123	1	2	3	0
User 3	0	0	123	1	2	3
User 4	3	0	0	123	1	2
User 5	2	3	0	0	123	1
User 6	1	2	3	0	0	123
User 7	123	1	2	3	0	0
User 8	0	123	1	2	3	0
User 9	0	0	123	1	2	3

Legend: 1 – Relevance feedback, 2 – Timed feedback, 3 – No feedback (None), 0 – News article not read.

**Table 4.2: Distribution of feedback methods per news and user.**

The tests were split into two parts:

- **Step 1** – Subjects were asked to give general information (age, gender and background);
- **Step 2** – Subjects read a total of six news articles, two with each of the feedback methods – Relevance Feedback, Timed Feedback and No Feedback. After watching each multimedia presentation the user had to rank the method;

To select a news article, the headline selection dialog box was modified with the introduction of the feedback options (Figure 4.3). An additional button was added to reset the user feedback (*Reset Feedback*). Test users were asked to use the button when changing news with different feedback settings. Not resetting the feedback changes the next reading of news. For example, two news articles with relevance feedback were read sequentially. Therefore, the second article has the first article’s feedback choices, which alters the presentation from the start. Using the reset feedback button will erase the illustration rates.

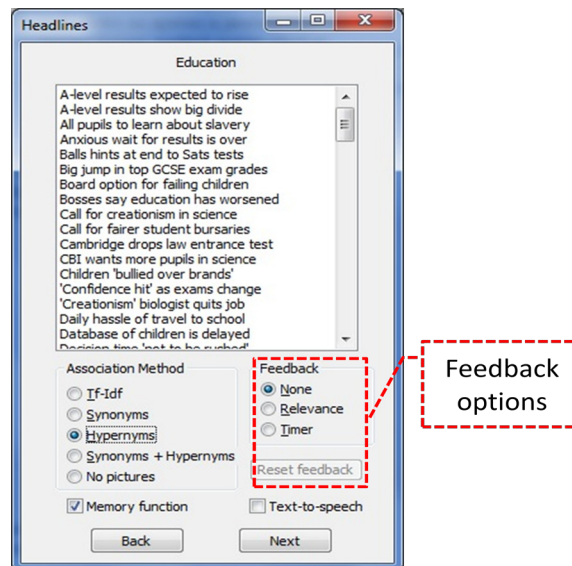


Figure 4.3: News Headline with highlight on the feedback options.

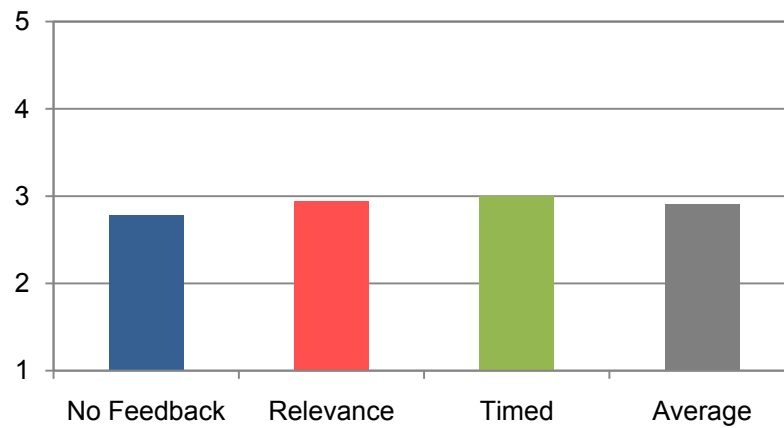
Users watched the news audio/visual presentation through the interface depicted in Figure 4.1. During the presentation, sentences and images were displayed. For each sentence/image combination, users had to mark the picture with one of the three classifications: *Like*, *Don't Like* and *Inadequate*. After the test, they were asked to give comments or suggestions regarding the application.

#### 4.4.2 Results and Discussion

Subjects ranked from 1 (worst) to 5 (best) each method after watching one article with the respective method (even after having rated the sentence-image associations, during the presentation). General assessment of the methods proposed discloses the Timed Feedback algorithm as the method with the best results. Table 4.3 shows the user feedback average results. As the table shows, an improvement is obtained from the usage of user feedback in relation to not using.

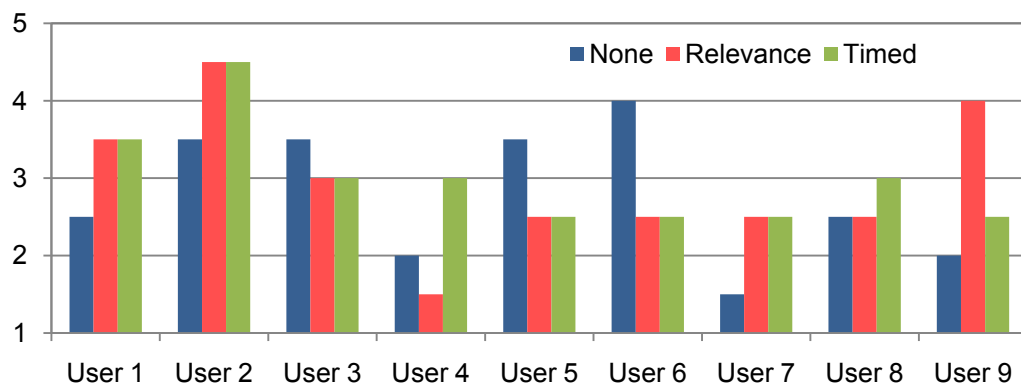
Feedback method	Rate
No Feedback	2.78
Relevance	2.94
Timed	3.00
Average	2.91

Table 4.3: General user assessment of the feedback methods.



**Figure 4.4: General user assessment of the feedback methods.**

For better understanding the obtained results, Figure 4.5 shows the general user assessment of the feedback methods per user. Recall that users read two news articles for each of the evaluated method. It is possible to see the scores discrepancy between different users which proves the ambiguity of the problem presented in this thesis. Different people are likely to have different preferences, even when evaluating photographs.



**Figure 4.5: Results for user feedback methods per news.**

User feedback techniques are not expected to give perfect results at the first trials. Results accuracy and precision are improved in the long-term. Figure 4.6 shows the average results for each feedback algorithm while discriminating both news articles read with the same algorithm in each evaluation. Results show a decline between first and second presentation with the same feedback option which is contrary to the expected results but can be explained by certain factors:

- A small number of test subjects. The lack of time and conditions to acquire users played a part in this issue;
- Techniques of long-term-expected results were used which makes it harder to obtain accurate results in short time;
- A small number of experiments were performed. Users tend to dislike long tests and for this reason only six news were read per evaluation;

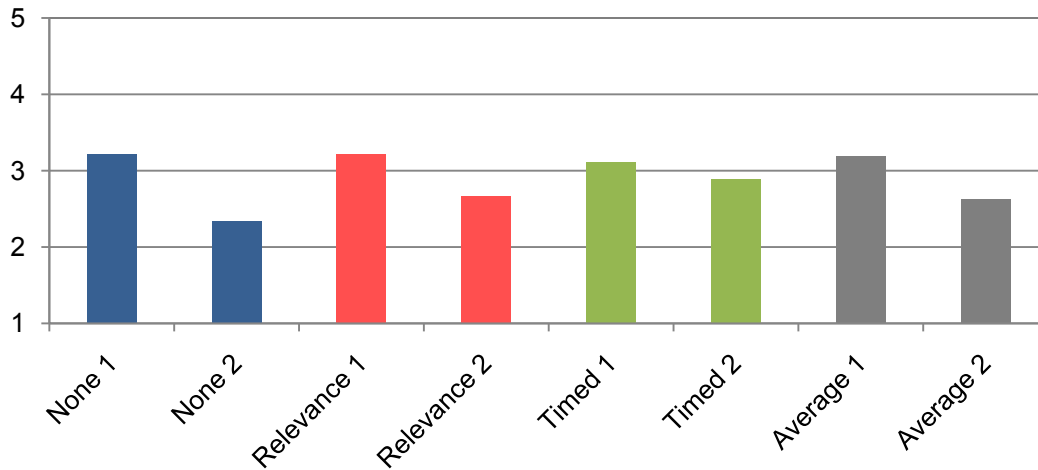


Figure 4.6: Assessment of the feedback methods divided by news read.

Even with the general user assessment of news, the user rates for the sentence-image associations were saved for statistical purposes. Figure 4.7 illustrate the times users pressed the feedback buttons during the presentations. Recall that users were asked to separate the *Like* and *Don't Like* from the *Inadequate* rate based on the adequacy between sentence and image.

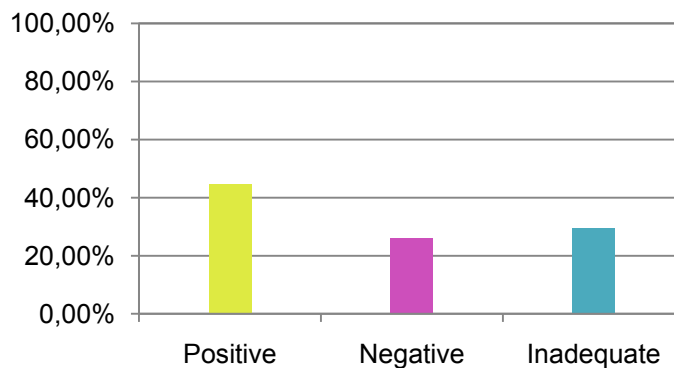


Figure 4.7: Feedback rates on the user evaluation.

## 4.5 Summary

In this chapter, an option for relevance feedback was introduced. Related work on RF systems were described and adapted to our framework. Two methods for gathering user feedback were proposed and tested properly by users. The user interface was modified – more precisely the News Headline and News reading modes – to support the user feedback options and buttons. The proposed method of explicit relevance feedback, adapted from the Rocchio formula, expects to receive illustration rates, given by users while at reading time. Users mark a sentence illustration depending on their preferences and in the next sentences, those preferences will influence the next illustrations. The second proposed method combines explicit feedback techniques with an implicit feedback element, inferred by the user reading time. Only illustration rates given at a usual user reading speed will produce a similar effect in the next illustrations, as the relevance feedback.

The development of the framework led to an improvement of the SSC method with the title and full news text becoming part of the similarity calculation.

Finally, user evaluations were conducted to assess the proposed feedback methods and results confirm an improvement in accuracy in relation to news read without user feedback.





## **Chapter 5**

# **Conclusions and Future Work**

Automated story illustration is not a new subject and several papers have been presented on this matter. In this thesis we proposed a framework that generates automated multimedia presentations to assist news readers. Our story illustration system is divided in three parts: user interface, sentence-image association algorithms and user feedback. From the researched work available, it is clear that more advances were made with textual analysis than with visual analysis which is understandable given the vast amount of textual information on the World Wide Web.

### **5.1 Conclusions and Contributions**

The framework's interface was carefully designed and its features described. The methods to compute the most suited image for each sentence can be divided in three parts: text analysis, term weighting and semantic similarity. An ontology was explored to refine the sentence-image relationship and a method to improve the coherence between sequentially selected illustrations was implemented (Story Sequence Consistency). A general user evaluation confirmed the hypothesis that illustrations are both useful and enjoyable. Results showed that hypernym methods with story sequence consistency were the best combination of sentence-image association.

Moreover, most users became actually interested in the application and provided us with valuable feedback and requests concerning their overall user

experience. For example, some suggested eliminating repeated illustrations in the same story (even if they are related to the story) or replacing the voice used for a less “synthetic” voice.

To filter non-relevant images that still illustrate sentences with unrelated content, user feedback functionalities were included in the framework. A relevance feedback method was implemented, based on the Rocchio algorithm, to separate relevant from non-relevant images. Another feedback method was proposed in this thesis: timed feedback. This method combines explicit feedback from the Rocchio algorithm with the concept of implicit “time-based” feedback: user reading time is taken into account when assigning image relevance. Again, user tests were performed to evaluate the proposed methods. Results provided evidence that feedback based personalization is a promising solution to automate the illustration of news text. In summary, the main contributions of this thesis are:

- An application to assist users in reading news;
- User evaluations which assessed the effectiveness of the overall framework and individual parts of the framework;
- An automatic sentence illustration algorithm was proposed, the Story Sequence Consistency. This algorithm achieved the best performance during the user evaluation;
- The user evaluation also showed that user feedback based illustration methods offer an improvement over automatic sentence illustration methods.

Finally, some of these contributions were published at international conferences:

- Diogo Delgado, João Magalhães, Nuno Correia, “Assisted news reading with automated illustration”, ACM Multimedia, technical demo, Florence, Italy, October 2010;
- Diogo Delgado, João Magalhães, Nuno Correia, “Automated illustration of news stories”, IEEE International Conference on Semantic Computing, Pittsburgh, United States of America, September 2010;

## 5.2 Future Work

### **Automatic Tag Correction and Annotation**

One of the problems encountered while developing the framework was the innumerable incorrectly spelled tags. A speller and a translator for the most common non-English languages would correct several tags.

Another limitation of the system was provided by the large number of poorly annotated images. For an image to be completely described, it would need a large set of multiple tags. To achieve a large number of tags per image, automatic image annotation algorithms could be used to compute the probability of new tags. Adding automatic tag correction and expansion would be a benefit to the framework.

### **User personalization**

The proposed feedback features provide the system with user preferences. However, one user's feedback can hinder other user's preferences. Some users may mark an illustration as non-relevant where others marked as relevant. Negative feedback marks reverse the effect produced by the same amount of positive marks. Adapting user personalization to the framework would enhance the user experience and result in more accurate illustrations. By saving the feedback rates, users can continue reading news with illustrations generated according to the previously given image rates.

### **User Attention based Feedback**

The main goal of the application is to draw the users' attention to the news content through the use of pictures to emphasize the text. While reading news, users look directly at the screen to see the illustrations and the news text. This setting allows the introduction of a visual analysis framework to detect users' reactions towards the news. This can be achieved using the OpenCV<sup>8</sup> library, comprising a set of programming functions for real time computer vision. The library is cross-platform and it can be adapted to the application.

Research in the area of user feedback has been primarily focused on user interaction data that is easily obtained using keyboard and mouse input. Actions such as click-through, display time, scrolling, and mouse movements are interactions that

---

<sup>8</sup> <http://opencv.willowgarage.com/wiki/>

can be easily measured. In several cases, these measures are used to estimate the relevance of entire documents. Recently, research has been conducted to clarify how other measures like eye movements (Buscher, Dengel et al. 2008; Buscher, Dengel et al. 2008) or emotions (Arapakis, Jose et al. 2008) could relate to user intent and user assessment of individually perceived relevance. Lee, Chang et al. (2007) attempt to improve user interaction by using a physical interface whereas we will use a camera to infer implicit user feedback.

The current method of illustration selection, the Story Sequence Consistency, uses a combination of the hypernyms based similarity together with a sliding window. The feedback functionality provides users with the possibility of guiding the selection of sentence illustrations. Introducing an emotion-based feedback can be translated into a modification in the ranking calculations. This means that at runtime, in the News Reading mode, illustrations will be rated accordingly to the user state.

In News Reader mode, we will use a camera to detect the users' attention when reading the news article. This will allow the application to detect the awareness/emotional state of the user and change the images and voice volume to engage the reader. Two states will be defined:

- **Attentive** – state where the user is looking at the monitor and reading the news presented; this state will provide positive feedback to the selected illustrations and news sections;
- **Unfocused** – state that can be defined as the act of not looking at the monitor while the text is shown; to serve as the worst case scenario, where the user dislikes the images being displayed; it will provide negative feedback to the selected illustrations and news sections;

The illustration selection algorithm will be combined with the user feedback provided by the emotion detection (attentive and unfocused), which means that a factor will be inferred from the user emotional state and incorporated in the image-sentence association.

We believe that these modifications will greatly improve the user experience.

## References

- Alborzi, H., A. Druin, et al. (2000). Designing StoryRooms: interactive storytelling spaces for children. Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques. New York City, New York, United States, ACM: 95-104.
- Arapakis, I., J. M. Jose, et al. (2008). Affective feedback: an investigation into the role of emotions in the information seeking process. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. Singapore, Singapore, ACM: 395-402.
- Balabanovic, M., L. L. Chu, et al. (2000). Storytelling with digital photographs. Proceedings of the SIGCHI conference on Human factors in computing systems. The Hague, The Netherlands, ACM: 564-571.
- Barker, K. and N. Cornacchia (2000). Using Noun Phrase Heads to Extract Document Keyphrases. Proceedings of the 13th Biennial Conference of the Canadian Society on Computational Studies of Intelligence: Advances in Artificial Intelligence, Springer-Verlag: 40-52.
- Barnard, K. and D. Forsyth (2001). Learning the semantics of words and pictures. Proceedings of the International Conference on Computer Vision.
- Berger, A. L. and V. O. Mittal (2000). OCELOT: a system for summarizing Web pages. Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. Athens, Greece, ACM: 144-151.

- Boguraev, B. and C. Kennedy (1999). "Applications of term identification technology: domain description and content characterisation." *Nat. Lang. Eng.* 5(1): 17-44.
- Brin, S. and L. Page (1998). "The anatomy of a large-scale hypertextual Web search engine." *Comput. Netw. ISDN Syst.* 30(1-7): 107-117.
- Buckley, C. and G. Salton (1995). Optimization of relevance feedback weights. Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval. Seattle, Washington, United States, ACM: 351-357.
- Buscher, G., A. Dengel, et al. (2008). Eye movements as implicit relevance feedback. CHI '08 extended abstracts on Human factors in computing systems. Florence, Italy, ACM: 2991-2996.
- Buscher, G., A. Dengel, et al. (2008). Query expansion using gaze-based feedback on the subdocument level. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. Singapore, Singapore, ACM: 387-394.
- Carson, C., S. Belongie, et al. (2002). "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying." *IEEE Trans. Pattern Anal. Mach. Intell.* 24(8): 1026-1038.
- Chu, M., Y. Li, et al. (2007). Enrich web applications with voice internet persona text-to-speech for anyone, anywhere. Proceedings of the 12th international conference on Human-computer interaction: intelligent multimodal interaction environments. Beijing, China, Springer-Verlag: 40-49.
- Claypool, M., P. Le, et al. (2001). Implicit interest indicators. Proceedings of the 6th international conference on Intelligent user interfaces. Santa Fe, New Mexico, United States, ACM: 33-40.
- Coyne, B. and R. Sproat (2001). WordsEye: an automatic text-to-scene conversion system. Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM: 487-496.
- Dagan, I., L. Lee, et al. (1999). "Similarity-Based Models of Word Cooccurrence Probabilities." *Mach. Learn.* 34(1-3): 43-69.
- Daille, B., É. Gaussier, et al. (1994). Towards automatic extraction of monolingual and bilingual terminology. Proceedings of the 15th conference on

- Computational linguistics - Volume 1. Kyoto, Japan, Association for Computational Linguistics: 515-521.
- Evans, D. K., J. L. Klavans, et al. (2000). Document processing with LinkIT. Proceedings of the RIAO Conference, Paris, France.
- Fox, S., K. Karnawat, et al. (2005). "Evaluating implicit measures to improve web search." *ACM Trans. Inf. Syst.* 23(2): 147-168.
- Goecks, J. and J. Shavlik (2000). Learning users' interests by unobtrusively observing their normal behavior. Proceedings of the 5th international conference on Intelligent user interfaces. New Orleans, Louisiana, United States, ACM: 129-132.
- Huiskes, M. J. and M. S. Lew (2008). The MIR flickr retrieval evaluation. Proceeding of the 1st ACM international conference on Multimedia information retrieval. Vancouver, British Columbia, Canada, ACM: 39-43.
- Hulth, A. (2003). Improved automatic keyword extraction given more linguistic knowledge. Proceedings of the 2003 conference on Empirical methods in natural language processing - Volume 10, Association for Computational Linguistics: 216-223.
- Jones, K. S. (1988). A statistical interpretation of term specificity and its application in retrieval. Document retrieval systems, Taylor Graham Publishing: 132-142.
- Joshi, D., J. Z. Wang, et al. (2006). "The Story Picturing Engine---a system for automatic text illustration." *ACM Trans. Multimedia Comput. Commun. Appl.* 2(1): 68-89.
- Justeson, J. S. and S. M. Katz (1995). "Technical terminology: some linguistic properties and an algorithm for identification in text." *Natural Language Engineering* 1(1): 9-27.
- Kageura, K. and B. Umino (1996). "Methods of automatic term recognition: a review." *Terminology* 3(2): 259.
- Kleinberg, J. M. (1999). "Authoritative sources in a hyperlinked environment." *J. ACM* 46(5): 604-632.
- Lee, C.-H. J., C. Chang, et al. (2007). Emotionally reactive television. Proceedings of the 12th international conference on Intelligent user interfaces. Honolulu, Hawaii, USA, ACM: 329-332.

- Li, L., Y. Shang, et al. (2002). Improvement of HITS-based algorithms on web documents. Proceedings of the 11th international conference on World Wide Web. Honolulu, Hawaii, USA, ACM: 527-535.
- Lovins, J. B. (1968). "Development of a stemming algorithm." *Mechanical Translation and Computational Linguistics* 11: 22-31.
- Lu, Y., C. Hu, et al. (2000). A unified framework for semantics and feature based relevance feedback in image retrieval systems. Proceedings of the eighth ACM international conference on Multimedia. Marina del Rey, California, United States, ACM: 31-37.
- Luhn, H. P. (1957). "A statistical approach to mechanized encoding and searching of literary information." *IBM J. Res. Dev.* 1(4): 309-317.
- Ma, W.-Y. and B. S. Manjunath (1999). "NeTra: a toolbox for navigating large image databases." *Multimedia Syst.* 7(3): 184-198.
- Mani, I. (1999). *Advances in Automatic Text Summarization*, MIT Press.
- Matsuo, Y. and M. Ishizuka (2004). "Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information." *International Journal on Artificial Intelligence Tools* 13(1): 157-169.
- Mayer, R., K. Steinhoff, et al. (1995). "A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text." *Educational Technology Research and Development* 43(1): 31-41.
- Noreault, T., M. McGill, et al. (1981). A performance evaluation of similarity measures, document term weighting schemes and representations in a Boolean environment. Proceedings of the 3rd annual ACM conference on Research and development in information retrieval. Cambridge, England, Butterworth & Co.: 57-76.
- Peng, J. (2003). "Multi-class relevance feedback content-based image retrieval." *Comput. Vis. Image Underst.* 90(1): 42-67.
- Pereira, F., N. Tishby, et al. (1993). *Distributional Clustering Of English Words*. Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics.
- Porter, M. F. (1980). "An algorithm for suffix stripping." *Readings in information retrieval* 14(3): 130-137.



- Salton, G. and M. J. McGill (1986). Introduction to Modern Information Retrieval, McGraw-Hill, Inc.
- Smeulders, A. W. M., M. Worring, et al. (2000). "Content-Based Image Retrieval at the End of the Early Years." IEEE Trans. Pattern Anal. Mach. Intell. 22(12): 1349-1380.
- Speer, N. K., J. R. Reynolds, et al. (2009). "Reading Stories Activates Neural Representations of Visual and Motor Experiences." Psychological Science 20(8): 989-999.
- Tanaka, K. and H. Iwasaki (1996). Extraction of lexical translations from non-aligned corpora. Proceedings of the 16th conference on Computational linguistics - Volume 2. Copenhagen, Denmark, Association for Computational Linguistics: 580-585.
- W. M. Shaw, J. (1995). "Term-relevance computations and perfect retrieval performance." Inf. Process. Manage. 31(4): 491-498.
- Wang, J. Z., J. Li, et al. (2001). "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIBraries." IEEE Trans. Pattern Anal. Mach. Intell. 23(9): 947-963.
- Wei, X., L. Yin, et al. (2004). Avatar-mediated face tracking and lip reading for human computer interaction. Proceedings of the 12th annual ACM international conference on Multimedia. New York, NY, USA, ACM: 500-503.
- White, R. W., J. M. Jose, et al. (2006). "An implicit feedback approach for interactive information retrieval." Inf. Process. Manage. 42(1): 166-190.
- Zigoris, P. and Y. Zhang (2006). Bayesian adaptive user profiling with explicit & implicit feedback. Proceedings of the 15th ACM international conference on Information and knowledge management. Arlington, Virginia, USA, ACM: 397-404.



# **Annexes**



# Assisted news reading with automated illustrations

Diogo Delgado, João Magalhães, Nuno Correia  
Department of Computer Science, Faculty of Science and Technology  
Universidade Nova de Lisboa, Portugal

[diogommdelgado@gmail.com](mailto:diogommdelgado@gmail.com), [imag@di.fct.unl.pt](mailto:imag@di.fct.unl.pt), [nmc@di.fct.unl.pt](mailto:nmc@di.fct.unl.pt)

## ABSTRACT

We all had the problem of forgetting about what we just read a few sentences before. This comes from the problem of attention and is more common with children and elderly. People feel either bored or distracted by something more interesting. This paper proposes an application to help people reading news by illustrating the news story. The application provides mechanisms to (1) select the best illustration for each scene and (2) to select the set of illustrations to improve the story sequence. The application proposed in this technical demo aims at improving the user's attention when reading news articles. The application implements several information processing techniques to generate an audio-visual presentation of the text news article.

## Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems], H.5.2 [User Interfaces]

## General Terms

Algorithms, Experimentation, Human Factors.

## Keywords

Automated illustration, text-to-image.

## 1. INTRODUCTION

One common use of pictures is to share stories and experiences with family and friends. Everyone at some point had the experience of browsing a photo album while the photographer tells us the story behind the pictures. Stories can be told with a set of photographs and a few comments added to the pictures – the human mind fills in the gaps. Automated story illustration may also be of use when in need for attention enhancement. Pictures often draw the readers' attention, so combining text with context based images is a positive way to engage elderly population in news reading, or even a younger person's willingness towards reading the current news. The goal of this paper is to demonstrate an application that generates an audio-visual presentation of a news article with image illustrations corresponding to the context of the story and the sentence being read, while responding to users' feedback. The following section presents the related work, section 3 describes our approach to the problem, section 4 gives a step-by-step description of the demo. Finally, section 5 discusses a user study.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'10, October 25–29, 2010, Firenze, Italy.  
Copyright 2010 ACM 978-1-60558-933-6/10/10...\$10.00.

## 2. AUTOMATED STORY ILLUSTRATION

The creation of multimedia stories by humans is a process that relies on the authors' imagination and on the media resources they produce and edit. Currently it is common to use computer software for this task as is the case of the work proposed by Balabanovic et al. [2]. They describe the implementation of a device that provides a convenient way of sharing digital photographs and associated stories with family and friends, in an easier way than the conventional album browsing.

Automatic text-to-scene conversion using computer graphics techniques has been studied for several years. One of such examples is StoryRooms [1], a physical interactive space for children created at the University of Maryland. Others have focused on understanding the text: WordsEye [3] is a system that parses natural language and converts it into three-dimensional scenes that represent the given text.

Other approaches, as is ours, use a repository of images to illustrate a given story. In [6], a technique for image ranking and selection based on a mutual reinforcement principle was presented. The set of candidate images is assumed to form a graph with the images acting as nodes and image similarities forming the weights on the edges.

## 3. ASSISTED NEWS READING

The assisted news reading application is composed by three dialog-based windows and a full screen mode: news category selection, news section, news headline selection and the news reading mode. The actual display of the news occurs after choosing the news, in the news reading screen, Figure 5. The time to display a sentence/paragraph is calculated based on the number of words. The key aspects of the system at the users' disposal are:

- **Text-to-speech.** This functionality is aimed at users who have eyesight problems or have reading difficulties.
- **Automated illustration method.** Images to illustrate a given news sentence are selected through the relation of the neighboring sentences and the image tags.
- **User feedback.** The user can mark the image as a good illustration or as a bad illustration.

Next, we shall describe the details of our approach.

### 3.1 News and images data processing

The news articles were collected from the BBC Web site and the images are from Flickr ([www.flickr.com](http://www.flickr.com)). The image dataset is composed by a 25,000 images and the corresponding tags [5]. Both news and tags were processed with standard techniques (punctuation removal, stop-words, stemming and tf-idf weighting). Both news data and pictures were stored in an SQLite database. Figure 2 illustrates how the data is processed by the application.

### 3.2 Selection of illustrations

Several methods were researched to associate images to a given sentence, see [4] for details. Since an article is presented sentence by sentence, we have to select an image for each sentence.

The standard method implements a cosine similarity between the current sentence and the candidate image tags. Previously, standard processing techniques were applied to the text. To expand the relation between images and sentences we incorporated WordNet [7] by adding synonyms to the image tags. This last change increased the similarity levels.

To improve the visual coherence of a sequence of illustrations we implemented a memory based function. With this method, the image-sentence selection procedure takes into account not only the current sentence but also the previous sentences. This technique is implemented as a sliding window based on the similarity between the previous sentence and image tags.

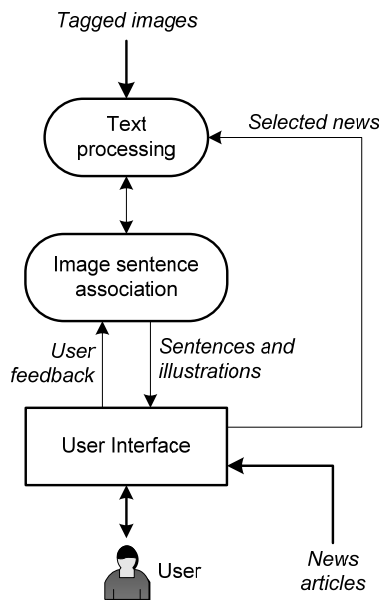


Figure 1: Application architecture.

Feedback was incorporated to offer users a certain degree of personalization and to improve the illustration selection method. This allows the system to ignore tags and to mark some images as inadequate for specific news. This is an important aspect as the system can infer the user’s interests concerning some news topics.

### 4. RUNNING THE DEMO

The application was developed as a Microsoft Windows application in C++ following a Model-View-Controller framework. All data is stored in a SQLite database. The application was developed for single-users.

To run the application the user simply has to start the application on any Windows based machine. Users start by selecting the subject of the news they intend to read about – dialog box of Figure 2. Once users have selected a subject, the next button will take them to the next dialog box. On the second dialog box, Figure 3, users select the section of the news articles. The next button will take them to headlines selection dialog box illustrated on Figure 4.

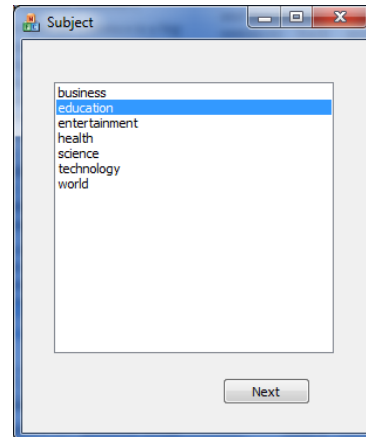


Figure 2: News subject selection.

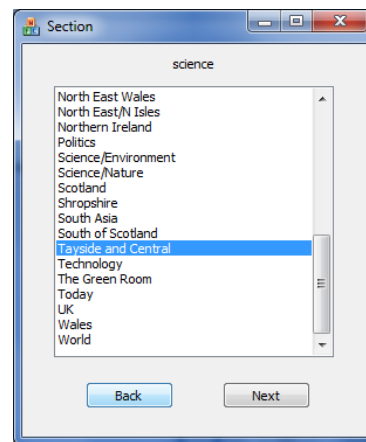


Figure 3: News section selection.

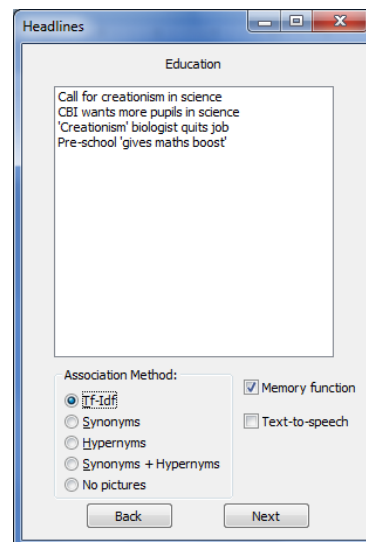


Figure 4. Selection of news headline and rendering options.

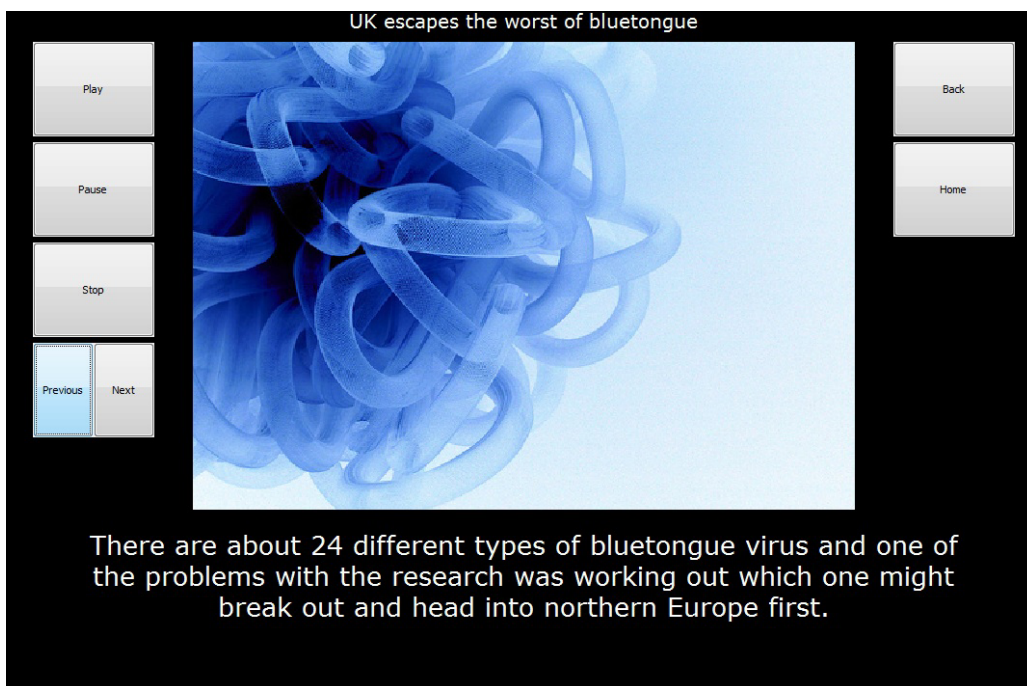


Figure 5. Examples of the full screen news reading mode.

On the third dialog box, Figure 4, users select not only the news headline but also the news illustration method, combining it with a memory based method and text-to-speech functionalities, see [4] for details. Once users finish this last dialog box, the next button starts the presentation of the news article.

The news article is rendered in full screen mode as is illustrated by Figure 5. News text is presented in a sentence by sentence fashion and each sentence is illustrated by automatically selected images. The news title is presented at the top of the screen and the current sentence is presented below the image. During the news presentation, illustrations are passed as a slide show with captions corresponding to sentences of the chosen news. To assist users, we provided a text-to-speech functionality to read the text out-loud. On the full screen view, the user can pause, resume or stop the news presentation (PAUSE button, PLAY button and STOP button respectively). The news can be played faster with the NEXT button or slower with the PREVIOUS button to jump between sentences.

Users can provide positive feedback with the RETURN or SPACE key and negative feedback with the DELETE key. Negative feedback is used to penalize both images and tags to prevent the application from selecting that same image for that particular news article.

Finally, users can return to the first dialog (HOME button) and select a different category/section/headline or go to the headlines dialog (BACK button) to select an article from the same section.

## 5. USER STUDY AND DISCUSSION

A set of experiments were conducted to assess the application: 18 subjects, aged from 22 to 43 years, both academics and graduate students in the area of computer science. The user study was performed on a Windows machine and all users had headphones to isolate ambient noise.

Each subject read six news articles. After reading the news they were asked to identify the application's most valuable features. Table 1 summarizes the means of the obtained results. In the last three questions, users were asked to quantify their opinion with a grade between 1 (worst) and 5 (best).

This user study suggests the application can indeed assist users in reading news. All users preferred the illustrations but, only 72.2% preferred audio which is somewhat surprising that subjects did not prefer audio content as much as visual content.

Question	Result
Do you prefer news with illustrations?	100% (YES)
Do you prefer news with sound?	72,22% (YES)
Were the illustrations useful?	3,83 / 5
Were the illustrations enjoyable?	3,83 / 5
Were the illustrations adequate?	3,50 / 5

Table 1. User study results.

The last three questions tried to assess if users found the illustrations useful for imagining the story, enjoyable because it makes the news more interesting, or if the selected illustration were not adequate. This last question provided us with an interesting result and feedback: users found some pictures to be related to the news content but not adequate to the news due to the

tone of the news. For example, if the tone of the news was sad or tragic, then an image portraying happiness or some positive emotion would not be adequate for that news article.

Finally, it should be noted that 100% of the 18 users preferred the news with illustrations but the last three questions suggests that the process of selecting the right illustrations should be improved.

**ACKNOWLEDGEMENTS.** This work has been partially funded by the *Fundação para a Ciência e Tecnologia*, Portugal, under project ARIA – Ambient-assisted Reading Interfaces for the Ageing-society, (PTDC/EIA-EIA/105305/2008). The authors would like thank the eighteen users that participated in the user evaluation.

## 6. REFERENCES

- [1] H. Alborzi, A. Druin, J. Montemayor, M. Platner, J. Porteous, L. Sherman, A. Boltman, G. Tax, J. Best, J. Hammer, A. Kruskal, A. Lal, T. P. Schwenn, L. Sumida, R. Wagner, and J. Hendler, "Designing StoryRooms: interactive storytelling spaces for children," Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques, New York City, New York, United States, 2000.
- [2] M. Balabanovic, L. L. Chu, and G. J. Wolff, "Storytelling with digital photographs," Proceedings of the SIGCHI conference on Human factors in computing systems, The Hague, The Netherlands, 2000.
- [3] B. Coyne and R. Sproat, "WordsEye: an automatic text-to-scene conversion system," Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 2001.
- [4] D. Delgado, J. Magalhaes, and N. Correia, "Automated illustration of news stories - Improving the readers experience," IEEE Intl Conference on Semantic Computing, Pittsburg, 2010.
- [5] M. J. Huiskes and M. S. Lew, "The MIR Flickr Retrieval Evaluation," ACM International Conference on Multimedia Information Retrieval Vancouver, Canada, 2008.
- [6] D. Joshi, J. Z. Wang, and J. Li, "The Story Picturing Engine---a system for automatic text illustration," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, pp. 68-89.
- [7] G. A. Miller, "WORDNET: A lexical database for English," *Communications of ACM*, vol. 38, pp. 39-41, November 1995.



# Automated illustration of news stories

## Improving the readers experience

Diogo Delgado, João Magalhães, Nuno Correia

Department of Computer Science, Faculty of Science and Technology

Universidade Nova de Lisboa

Lisbon, Portugal

diogomdelgado@gmail.com, jmag@di.fct.unl.pt, nmc@di.fct.unl.pt

**Abstract** — Forgetting what one has just read is, in some cases, linked to insufficient attention. The reader might feel either bored or distracted by something more interesting – a common trace in children and the elderly. The challenge is: how can multimedia systems assist readers in reading and remembering stories? Several studies [2, 7, 14] showed that reading memory is improved by visual stimulus. In this paper we formulate the hypothesis that an automated multimedia system can help users in reading a story by stimulating their reading memory with adequate visual illustrations. These illustrations are intended to increase the readers’ attention towards the story and to help them recalling the story. The framework automatically creates a multimedia presentation of the news story by (1) rendering the news text in a sentence-by-sentence fashion, (2) providing mechanisms to select the best illustration for each sentence and (3) select the set of illustrations that guarantees the best sequence of illustrations. Users may also activate a text-to-speech functionality according to their preference or reading difficulties. Experiments used Flickr images to illustrate BBC news articles: a user survey with 23 users assessed positively the effectiveness of the system.

**Keywords:** Story Illustration, Information Retrieval, Text-to-speech, Keyword Classification

### I. INTRODUCTION

When reading a book it is common for readers to create images in their minds depicting the written scene. For example, in Mary Shelly’s *Frankenstein*, the monster image depicted in our heads may differ from person to person. Sometimes, in an attempt to spark readers’ imagination, book illustrations are used to better engage the reader in the story [20]. Over the years computer software has been widely used by authors to create visual illustrations [7]. In this paper we study the hypothesis that computers can automatically select good visual illustrations from a large repository of photographs.

One common use for photographs is to share stories, experiences and holidays with family and friends. Everyone at some point had the experience of browsing a photo album while the photographer describes the story behind each picture. The World Wide Web facilitates the sharing of digital photos and has fostered the process of “digital storytelling”. Nowadays, there is no need to send photographs by e-mail since it is possible to upload pictures to a social network website such as Facebook or Flickr providing access to everyone involved in the album and more if it is the uploader’s desire. The story is told with a set of photographs and a few added comments – the human mind fills in the gaps. It is well

discussed by Speer et al. [20] who registered human brain activity while reading text stories. Reading about different visual scenes and motor experiences would activate particular parts of the brain.

In this setting the Web supply us with millions of tagged images, which can be useful to illustrate stories with real images that will capture readers’ curiosity and imagination. Searching for a video on YouTube or a photograph on Flickr requires users to describe the most important aspects of what they wish to find. Usually we don’t use full sentences as “Video where Mary went to a night festival in London”, but words like “Mary”, “festival” or “London”. These words are addressed as keywords. They define the sentences’ content and make search engines more effective. In our work, we make use of text keywords to search picture tags – user supplied text annotations.

#### A. Motivation

There are many areas where automatic creation of multimedia presentations from text can be of great value, e.g. entertainment, media, journalism, children’s picture books. For example, when journalists write news stories, they need to select images to illustrate parts of the story. These illustrations are usually chosen from a repository or from specific photos taken at the news setting. Automated story illustration may also be of use when in need for attention enhancement. Pictures often improve the readers’ attention: Mayers et al., [16], performed an experiment where students learning a subject from an only-text explanation performed worse than students who learned from illustrated text explanations.

#### B. Our Approach

Our main contribution is a framework to assist news readers by generating an audio/visual presentation of news to improve their attention. This presentation is generated with small parts of the original text and with the appropriate image illustration framed in the context of the last sentences. The key aspect of our contribution is a method to perform a semantic comparison between a range of sentences and image tags. We address this challenge in two ways. First, we expand sentences with a linguistic ontology (WordNet) to enclose all possible linguistic meanings. Second, because reading is a sequential process, we guarantee the sequential consistency of the chosen illustrations by selecting a set of semantically coherent images.

Next we revise the related work literature. In section III, we detail the assisted news reading framework in terms of its main components and formalize the problem. In Section IV we detail

the process of selecting semantic illustrations to generate the audio/visual presentation of news articles. In Section V we present the user tests and discuss the obtained results. In section VI, we summarize the main aspects of this paper and discuss future work.

## II. RELATED WORK

Story illustration is a research area involving expertise from different areas such as image retrieval, information retrieval and natural language processing. It addresses the problem of finding the best set of pictures to describe a piece of text. In the context of automated story illustration, our approach attempts to quantify images' importance, based on their annotations, to illustrate a given text.

### A. Automated illustration

The creation of stories by humans is a process that relies on the authors' imagination and on the resources they are creating and compiling. It is common nowadays to use computer software for this task as is the case of the work proposed by Balabanovic et al. [2]. They discuss the implementation of a device that provides a convenient way of sharing digital photographs and associated stories with family and friends, in an easier way than the conventional album browsing. While Balabanovic proposed a system to help humans assembling a multimedia story, in this paper we aim at researching a multimedia system that replace humans in this task by taking news stories and linking meaningful image illustrations to each news segment. Automatic text-to-scene conversion using computer graphics techniques is a different type of approach that has been researched. One of such examples is the StoryRoom [1], a physical interactive space for children built at the University of Maryland. Others have focused on understanding the text: WordsEye system developed at AT&T Labs [7] is a system that parses natural language and converts it into three-dimensional scenes that represent the given text. The goals of the former and the latter are very similar.

### B. Story text analysis

Inferring the elements from a text corpus that should be illustrated is a key part of our framework. It is vital to quantify the importance of each text term. In the information retrieval research field, term weighting is an important technique for document retrieval, search engines, document summarization, text mining, etc [13, 17]. The aim of keyword extraction is to find a small set of terms that describes a specific document, independently of the domain it belongs to. Hulth et al. [12] discussed an approach on automatic extraction of keyword terms using a supervised machine learning algorithm. Hulth came to the conclusion that by adding linguistic knowledge to the representation rather than relying only on statistics returns better results when compared to keyword tags. Since our method makes use of a large corpus, we prefer to use proved methods for term weighting and perform a semantic expansion

with a linguistic ontology and a method to guarantee a sequential coherence of the illustrations.

When part-of-speech (PoS) patterns are used to extract potential terms, the problem lies in how to restrict their number and keep only the relevant terms. Another problem happens when terms have more than one sense or represent more than one PoS at different times. Finding potential terms when no machine learning is involved in the process and by means of PoS patterns is a common approach. For example, Barker et al. [3] discusses an algorithm where the number of words and the frequency of noun phrases and their head noun are used to determine which terms are keywords. An extraction system called LinkIT [10] compiles the phrases having a noun as the head, and rank these according to the heads' frequency. This is similar to our combination of the stemming algorithm [18] with the weighting technique, but we quantify every text term and image tag. Boguraev et al. [5] extracted technical terms based on the noun phrase patterns suggested by Justeson et al. [15]; these terms are then the basis for a headline-like characterization of a document. The study carried out by Daille et al. [8] investigated the performance of statistical filters applied on extracted noun phrases and concluded that term frequency is the best candidate.

### C. Illustration analysis and selection

In the area of information retrieval this problem has been approached as an image ranking and selection problem: "how to choose the best set of images from an image database to illustrate a piece of text". Several efficient image retrieval systems have appeared in the last decade [6, 19, 21]. Some focus on quantifying image similarity – returning images similar to one image that serves as query. In Barnard et al. [4] the idea of auto-illustration as an inverse problem is introduced. Statistical associations between images and text were used to find images with high likelihood, given a piece of text. Computing the degree of association between blocks of information is a difficult task – mutual reinforcement principle-based methods have been widely reported. Joshi et al. [14], proposed a technique for image ranking and selection based on a mutual reinforcement principle and a discrete Markov chain model. The set of candidate images is assumed to form a graph with the images acting as nodes and image similarities forming the weights of the edges. Under a special condition, the image selection can be modeled as a random walk in the graph. However, even with a well annotated image database available, choosing a few images which best represent the text is not a simple task. The problem is subjective, and a human would have to make an empirical assessment, using knowledge gained over the years, to judge the significance of each image. Our framework follows this type of approach. However, we propose a finer grain approach aiming at illustrating individual sentences and not full stories.

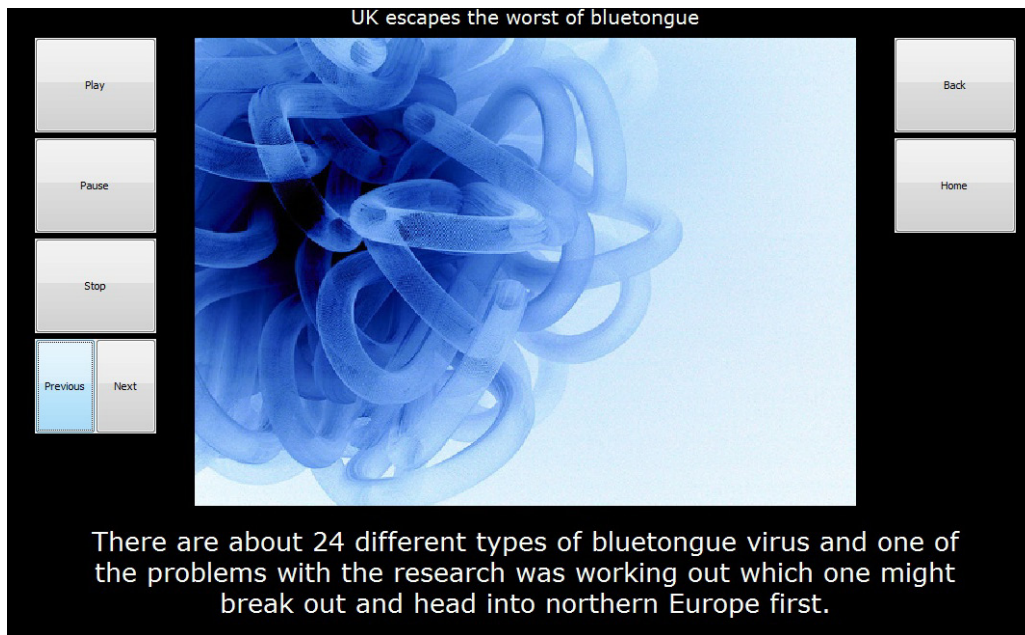


Figure 1: News reading mode.

### III. ASSISTED NEWS READING

The assisted news reading application is composed by three dialog-based windows and a full screen mode: news subject, section and headline selection and a news reading mode. The actual display of news occurs in the news reading screen after selecting a news article and illustration options. The key aspects of the system are a text-to-speech functionality, aimed at users who have eyesight problems or reading difficulties, and most importantly, an automated illustration selection method to illustrate news articles. This last method is based on the relationship between image tags and news terms.

A news article is rendered in full screen mode as is illustrated in Figure 1. The text is presented in a sentence by sentence fashion where each sentence is illustrated with automatically selected images. A sentence display time is calculated based on the number of words. The title is presented at the top of the screen and the sentences are presented below the image. During a news presentation, illustrations are passed as a slideshow with captions corresponding to sentences of the chosen news. On the full screen view, users can pause, resume or stop the news presentation. Also, news readers can jump between sentences with the NEXT or the PREVIOUS buttons. A demo paper was produced describing the assisted news reading application and its features [9].

The process to compute the similarity between news text and image tags is done in three steps: first we perform standard text processing techniques such as sentence extraction, stop-word removal and stemming; second, we weight each image annotation and news term with a weighting technique to quantify its importance; and thirdly, we compute the similarity between weighted image annotations and news terms. This baseline approach is then extended with an ontology-based expansion and a method to improve the coherence between sequential illustrations selections.

#### A. Text Processing

First, news articles were segmented into sentences, stop-words were removed and text terms were stemmed using the Porter stemmer algorithm. In linguistic morphology, stemming is a process for reducing words to their base, stem or root form. This algorithm was also performed on the image annotations to remove words of the same family.

#### B. Terms and Tags Weighting

In the second step, both terms and tags were weighted according to their frequency in the news article or image annotation set and to their frequency in the collection of news or images. For this purpose we applied the weighting technique term frequency-inverse document frequency (*tf-idf*). The *tf-idf* weight is a statistical measure to evaluate how important a word is to a document in a collection. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the collection of documents. After this first step, the  $n^{th}$  sentence of the  $m^{th}$  news article is represented by the vector

$$s_{m,n} = (t_1, \dots, t_T),$$

where each component  $t_p$  indicates the weight of the respective term, from a total of  $T$  news text terms.

Normally, *tf-idf* is only applied to text documents but in our case we need to consider the annotations importance. Thus, we also applied the *tf-idf* technique to tags in the images dataset. Moreover, since the weighting technique would favor rare (or misspelled) words combined with sets of few elements (images tags), we discarded images with less than 5 tags and tags with less than 10 occurrences. Thus, the  $k^{th}$  image is represented by the vector

$$i_k = (t_1, \dots, t_M),$$

where component  $t_p$  indicates the weight of respective tag, from a total of  $M$  tags.

### C. Sentence-Image Comparison

In the third step we compute the degree of association between an image and a sentence, taking into account weighted image-tags and sentence-terms. Note that in this step we are looking for the image that best matches a particular sentence. Thus, we need to rank images according to their relation to the sentence at hand. An image rank is computed with the cosine similarity (or cosine distance) which measures the angle between two vectors. In this case, the vectors represent sentences and images:

$$D_{\text{cosine}}(s_{m,n}, i_k) = \frac{s_{m,n} \times i_k}{\|s_{m,n}\| \times \|i_k\|}$$

where  $s_{m,n}$  is the  $n^{\text{th}}$  sentence of the  $m^{\text{th}}$  news article, and  $i_k$  is  $k^{\text{th}}$  image. With this method, it is possible to estimate the degree of relation between an image and a sentence. One problem with this process is the fact that only images that have the exact text words can be selected for illustration. A solution to this problem is presented in the next chapter.

## IV. SEMANTIC ILLUSTRATIONS

The defined framework for comparing news text to image tags is limited to their terms and tags, leaving out the true semantic interpretation of both elements. We tackle this challenge with two methods. First, we expand sentences and image tags with a linguistic ontology (WordNet) to enclose all possible meanings. Second, because reading is a sequential process, we attempt to construct a sequential consistency between selected images to guarantee a semantically coherent set of illustrations.

### A. Semantic Expansion

WordNet was used to create a group of synonyms and hypernyms – words whose semantic range includes the word being used (e.g. *animal* is hypernym of *dog*). This process is applied only to news nouns. For every noun  $t_i$  in the sentence vector  $s_{m,n} = (t_1, \dots, t_T)$ , we use WordNet to compute its synonyms,

$$\text{Syn}(t_i) = (ts_{i,1}, \dots, ts_{i,T}),$$

where each  $ts_{i,k}$  corresponds to a synonym of a term  $t_i$ . Synonyms will have the same weight as the original term. The resulting sentence is

$$s_{m,n} = (t_1, \dots, t_T) + \sum_{i \in \text{nouns}(t_1, \dots, t_T)} (ts_{i,1}, \dots, ts_{i,T}),$$

where the first part of the equation represents the original sentence and the second represents the sentence-nouns synonyms.

A similar result is obtained using the hypernyms function, where one word becomes a vector

$$\text{Hyp}(t_i) = (th_{i,1}, \dots, th_{i,T}),$$

where each component  $th_{i,k}$  corresponds to a hypernym of  $t_i$ . Again, hypernyms will have the same weight as the original term, which is also included in the sentence vector. When expanding hypernyms, the root of the concept hierarchy of a noun corresponds to the word “entity” or “abstract”. This would produce similar results for very different sentences. We propose a resolution to this predicament: If a noun is  $N$  hypernyms from the root we only consider the first  $N/2$  hierarchy levels; for example, the word “continent” hierarchical tree: continent  $\rightarrow$  landmass  $\rightarrow$  land  $\rightarrow$  object  $\rightarrow$  physical entity  $\rightarrow$  entity – would be transformed in (continent, landmass, land). The resulting sentence is

$$s_{m,n} = (t_1, \dots, t_T) + \sum_{i \in \text{nouns}(t_1, \dots, t_T)} (th_{i,1}, \dots, th_{i,T}),$$

where the first part of the equation represent the original sentence and the second represent the sentence-nouns hypernyms.

A third method to perform the semantic expansion is to consider both synonyms and hypernyms. Thus, a sentence is then represented by its terms and the synonyms and hypernyms of its nouns:

$$s_{m,n} = (t_1, \dots, t_T) + \sum_{i \in \text{nouns}(t_1, \dots, t_T)} [(ts_{i,1}, \dots, ts_{i,T}) + (th_{i,1}, \dots, th_{i,T})],$$

where the first part of the equation represent the original sentence and the second represent the sentence-nouns synonyms and hypernyms.

### B. Story Sequence Consistency (SSC)

To guarantee that selected illustrations have a similarity not only with one sentence, but with the content of the news article as well, we employ a memory based function to consider the previous sentences. A sliding window parses neighboring sentences to calculate the accumulated weight of each news term. The final ranking function

$$\text{Rank}(s_{m,n}, i_k) = \sum_{p=n-\text{WinSize}}^n \left[ \frac{1}{n-p+1} D_{\text{Cosine}}(s_{m,p}, i_k) \right],$$

computes the rank position of image  $k$  in relation with the  $n^{\text{th}}$  sentence of the  $m^{\text{th}}$  news article. The variable *WinSize* indicates the windows range to include the previous sentences. The factor  $(n-p+1)^{-1}$  is a weight decay to adjust the contribution of sentences according to their distance to the  $n^{\text{th}}$  sentence.

## V. EVALUATION

To evaluate our approach we conducted a user evaluation to assess the different illustration methods. For this purpose we used a dataset of BBC news downloaded from their website at <http://www.bbc.co.uk/> and a 25,000 image dataset obtained from Huskies et al. [11].

### A. BBC Web news Dataset

News articles were collected from the BBC website and are available for download<sup>1</sup>. A total of 6,727 news articles were collected with each belonging to just one subject and section. On the BBC website news URLs are organised according to subject. Therefore the news subject is available for extraction from the article’s URL. Each news section is obtained via assignment by BBC journalists.

Since the files from BBC are in *.html* format, a parser was used to extract the needed information. A SAX parser was developed to remove navigational content and extract the news corpus and title. Special care was taken to handle language specific characters and other formatting data. The resulting data was stored in a SQLite database.

### B. Flickr Images dataset

A collection comprising 25,000 Flickr<sup>2</sup> images was redistributed by Huskies et al. for research purposes. They extracted the image tags and other image metadata. The tags were inserted by Flickr users with a folksonomy which in some cases, results in insufficiently described images. From the whole dataset, 19,892 had at least five tags; thus, the remaining images were not used in our experiments.

### C. Experiment

To assess the general approach and the specific methods we conducted a user evaluation with 23 subjects. The user study was performed on a Windows machine and all users had headphones to isolate ambient noise. The tests were divided into three parts:

- **Step 1:** Subjects were asked to give general information (age, gender and background);
- **Step 2:** Subjects read six news articles, each with a combination of a specific illustration method with the SSC feature on or off. After watching each multimedia presentation the tester had to grade the method
- **Step 3:** Subjects responded to general questions about the application and were shown an illustration photograph to verify how much of the original news they recalled based on just that illustration.

### D. Results

#### 1) Illustration Method assessment

Users watched the news articles audio/visual presentation through the interface depicted in Figure 1. They were asked to grade from 1 (worst) to 5 (best) each method after watching a news article.

The story sequence consistency method presented better results than using a single sentence – SSC with hypernyms was the best method. With synonyms, the semantic expansion is limited to the same objects; consequently the sentence-image relation is not greatly improved. Alternatively, using hypernyms actually widens the semantic meaning of a sentence, capturing richer sentence-image relationships. In some cases, a few of the selected illustrations might be unrelated to the news content and the connection with the text can be hard to understand. This happens in some situations due

to external factors as improper image tagging. Using both synonyms and hypernyms should have improved the quality of the selected images, but instead, as we can observe in Table I, results show the opposite. This is related to a too wide semantic expansion of the initial nouns.

TABLE I. RESULTS OF THE ILLUSTRATION METHODS ASSESSMENT.

Illustration method	SSC off	SSC on
Synonyms	2,45/5	2,88/5
Hypernyms	2,38/5	2,94/5
Synonyms + Hypernyms	2,56/5	2,25/5

#### 2) General assessment

The general assessment conducted at the end of the experiment measured the effectiveness and viability of the proposed framework. Results suggest the application can indeed assist users in reading news. All test users preferred reading with the illustrations but, only 72.2% preferred audio in addition to the illustrations. It was somewhat surprising to see that some subjects preferred not to have the news presented orally.

TABLE II. RESULTS OF GENERAL ASSESSMENT EXERCISE.

Question	Result
Were the illustrations useful?	3,83 / 5
Were the illustrations enjoyable?	3,83 / 5
Were the illustrations adequate?	3,50 / 5

The purpose of the three questions in Table II was to assess the users’ impression concerning illustrations usefulness for imagining the story, “enjoyability” for making the news more interesting and compelling, and if the generally selected illustrations were adequate to the news context. This last question provided us with an interesting result and feedback: users found pictures to be related to the news content but some were not adequate due to the news tone. For example, if a news nature was sad or tragic, then an image portraying happiness or some positive emotion would not be adequate for that news.

TABLE III. RESULTS OF USER VISUAL MEMORY.

Recalls news title/content?	Result
No	11,1%
Vaguely	22,2%
Yes	66,7%

Finally, Table III presents the results concerning users’ visual memory. After reading six news articles users were shown pictures and were asked to describe the news it illustrated. The large majority answered correctly and recalled the full news title which suggests that illustrations were useful.

## VI. CONCLUSIONS

In this paper we proposed a framework that generates automated multimedia presentations to assist news readers. The method to compute the best illustrations for each sentence was carefully described. Also, an ontology was explored to refine the sentence-image relationship and a method to improve the

<sup>1</sup><http://ctp.di.fct.unl.pt/~jmag/BBCnewsDataset/>  
<sup>2</sup><http://www.flickr.com/>

coherence between sequentially selected illustrations was implemented. Results show that hypernyms with story sequence consistency were the best sentence-image association method. General user feedback confirmed the hypothesis that illustrations are both useful and enjoyable. Moreover, most users became actually interested in the application and provided us with valuable feedback and requests concerning their overall user experience. For example, some suggested eliminating repeated illustrations in the same story (even if they are related to the story) or replacing the voice used for a less “synthetic” voice. As future work we plan to improve the correlation between images in the same sequence to create a stronger sense of animated story.

#### ACKNOWLEDGEMENTS

This work has been partially funded by the *Fundação para a Ciência e Tecnologia*, Portugal, under project ARIA – Ambient-assisted Reading Interfaces for the Ageing-society, (PTDC/EIA-EIA/105305/2008). The authors would like thank the twenty three users that participated in the user evaluation.

#### REFERENCES

- [1] H. Alborzi, *et al.*, "Designing StoryRooms: interactive storytelling spaces for children," presented at the Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques, New York City, New York, United States, 2000.
- [2] M. Balabanovic, *et al.*, "Storytelling with digital photographs," presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, The Hague, The Netherlands, 2000.
- [3] K. Barker and N. Cornacchia, "Using Noun Phrase Heads to Extract Document Keyphrases," presented at the Proceedings of the 13th Biennial Conference of the Canadian Society on Computational Studies of Intelligence: Advances in Artificial Intelligence, 2000.
- [4] K. Barnard and D. Forsyth, "Learning the semantics of words and pictures," in *Proceedings of the International Conference on Computer Vision*, 2001, pp. 408-415.
- [5] B. Boguraev and C. Kennedy, "Applications of term identification technology: domain description and content characterisation," *Nat. Lang. Eng.*, vol. 5, pp. 17-44, 1999.
- [6] C. Carson, *et al.*, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 1026-1038, 2002.
- [7] B. Coyne and R. Sproat, "WordsEye: an automatic text-to-scene conversion system," presented at the Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 2001.
- [8] B. Daille, *et al.*, "Towards automatic extraction of monolingual and bilingual terminology," presented at the Proceedings of the 15th conference on Computational linguistics - Volume 1, Kyoto, Japan, 1994.
- [9] D. Delgado, *et al.*, "Assisted news reading with automated illustration," presented at the Proceedings of ACM Multimedia 2010 Florence, Italy, 2010.
- [10] D. K. Evans, *et al.*, "Document processing with LinkIT," in *Proceedings of the RIAO Conference*, Paris, France, 2000.
- [11] M. J. Huiskes and M. S. Lew, "The MIR flickr retrieval evaluation," presented at the Proceeding of the 1st ACM international conference on Multimedia information retrieval, Vancouver, British Columbia, Canada, 2008.
- [12] A. Hulth, "Improved automatic keyword extraction given more linguistic knowledge," presented at the Proceedings of the 2003 conference on Empirical methods in natural language processing - Volume 10, 2003.
- [13] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," in *Document retrieval systems*, ed: Taylor Graham Publishing, 1988, pp. 132-142.
- [14] D. Joshi, *et al.*, "The Story Picturing Engine---a system for automatic text illustration," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, pp. 68-89, 2006.
- [15] J. S. Justeson and S. M. Katz, "Technical terminology: some linguistic properties and an algorithm for identification in text," *Natural Language Engineering*, vol. 1, pp. 9-27, 1995.
- [16] R. Mayer, *et al.*, "A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text," *Educational Technology Research and Development*, vol. 43, pp. 31-41, 1995.
- [17] T. Noreault, *et al.*, "A performance evaluation of similarity measures, document term weighting schemes and representations in a Boolean environment," presented at the Proceedings of the 3rd annual ACM conference on Research and development in information retrieval, Cambridge, England, 1981.
- [18] M. F. Porter, "An algorithm for suffix stripping," in *Readings in information retrieval*, ed: Morgan Kaufmann Publishers Inc., 1997, pp. 313-316.
- [19] A. W. M. Smeulders, *et al.*, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 1349-1380, 2000.
- [20] N. K. Speer, *et al.*, "Reading Stories Activates Neural Representations of Visual and Motor Experiences," *Psychological Science*, vol. 20, pp. 989-999, August 2009 2009.
- [21] J. Z. Wang, *et al.*, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 947-963, 2001.

# Story illustrator & Illustration Methods Assessment

**Demographics:**

How old are you? \_\_\_\_\_

What is your gender? [Male] [Female]

What is your profession? \_\_\_\_\_

Do you consider yourself a “computer literate”? [Yes] [No]

Date: \_\_\_\_\_

**Start the application and view the following news with the specified options:**

Subject	Section	Headline	Option	Text-to-Speech	Memory function	Grade from 1 to 5 the adequacy between pictures and news
Science	England	Uk escapes the worst case of bluetongue	No pictures	Yes	No	
Science	England	Uk escapes the worst case of bluetongue	Tf-idf	Yes	No	
World	Science/Nature	Ancient trees recorded in mines	Synonyms	No	No	
Health	Education	Free cook books for 11-year old	Tf-idf	Yes	Yes	
World	Business	Boeing workers decide to strike	Hypernyms	Yes	Yes	
Science	Europe	Belgians warned over iodine leak	Hypernyms + Synonyms	Yes	No	

**After watching some text and illustrated news:**

How often do you read news on the internet? [Never] [Sometimes] [Weekly] [Daily]

Were the illustrations useful for imagining the news setting? [--] [-] [0] [+] [++]

Were the illustrations adequate for the news content? [--] [-] [0] [+] [++]

Did you enjoy the selected illustrations? [--] [-] [0] [+] [++]

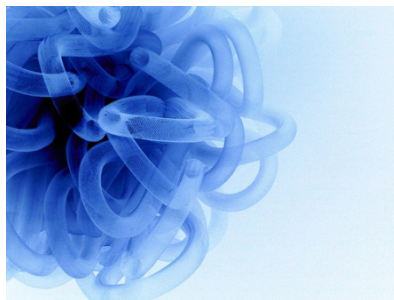
**After 10 mins:**

Which news do you recall best (title)? \_\_\_\_\_

Did you prefer the news with or without illustrations? [with] [without]

Did you prefer the news with or without sound? [with] [without]

**Show a picture and verify if readers remember the details of the story.**



Do you remember the title? [Yes] [No]

If yes, can you briefly describe the content of the story?

**Notes and suggestions:**



## User Feedback Assessment

**Demographics:**

How old are you? \_\_\_\_\_

What is your gender? [Male] [Female]

What is your profession? \_\_\_\_\_

Do you consider yourself a “computer literate”? [Yes] [No]

Date: \_\_\_\_\_

**Start the application and view the following news with the specified options:**

Subject	Section	Headline	Feedback option	Grade from 1 to 5 the adequacy between pictures and news
World	Asia-Pacific	Police clash with Thai protesters	None	
Health	Merseyside	Teddy camera catches career thief	None	
World	Asia-Pacific	Police clash with Thai protesters	Relevance	
Business	Hampshire	Ford bosses in talks with union	Relevance	
World	Asia-Pacific	Police clash with Thai protesters	Timer	
Science	Gloucestershire	Bid for world land-speed record	Timer	

**Notes and suggestions:**