

PERSONALISING THE EDITORIAL MIX FOR A DIGITAL NEWSPAPER USING CONSTRAINT PROGRAMMING

MICHAEL LUNØE



Department of Informatics and Mathematical Modelling
Technical University of Denmark

August 2012

Michael Lunøe: *Personalising the Editorial Mix for a Digital News-paper using Constraint Programming*, © August 2012

SUPERVISORS:

Michael Kai Petersen
Carsten Witt

ABSTRACT

Short summary of the contents...

CONTENTS

I EXPLORING THE FIELD	1
1 CREATING THE EDITORIAL MIX	3
1.1 The Editorial Mix Problem	3
1.2 Personalising a Digital Newspaper	5
1.3 Contributions of Constraint Programming	7
1.4 Problem Description	9
2 PERSPECTIVES AND INSPIRATIONS	11
3 FEATURES TO BE PERSONALISED	15
3.1 User Needs	15
3.2 Use Cases	18
3.3 Requirements	20
3.4 The Editorial Mix	24
II USING WHAT WE HAVE LEARNT	29
4 PROPOSED DESIGN	31
4.1 Layout and Typography	31
4.2 Interactions	32
4.3 Problem Representation	32
4.4 Delimitation	35
5 TECHNICAL SPECIFICATION	37
5.1 Constraint Personalisation Library	37
5.2 Front End	37
5.3 Backend worker.js	37
5.4 Server for Acquiring and Mining Data for Personalisation	40
III LEARNING WHAT WE HAVE USED	43
6 DISCUSSION	45
7 EVALUATION OF THE SOLUTION	47
7.1 Test	48
8 CONCLUSION	51
IV APPENDIX	53
A USER NEEDS	55
A.1 Personas	55
A.2 Scenarios	58
A.3 Business Case	61

A.4 Requirements	62
A.5 Test Results	65

BIBLIOGRAPHY	67
--------------	----

ACRONYMS

CP Constraint Programming Constraint programming is a programming paradigm wherein relations between variables are stated in the form of constraints.

CSP Constraint Satisfaction Problem Mathematical problems defined as a set of objects whose state must satisfy a number of constraints or limitations.

COP Constraint Optimisation Problem Can be defined as a regular constraint satisfaction problem in which constraints are weighted and the goal is to find a solution maximizing the weight of satisfied constraints.

RSS Really Simple Syndication. A family of web feed formats used to publish frequently updated works—such as blog entries, news headlines, audio, and video—in a standardized format.

Part I

EXPLORING THE FIELD

Text for part 1.

1 | CREATING THE EDITORIAL MIX

This chapter introduces the editorial mix of a digital newspaper and which parameters to account for when composing the newspaper. It is afterwards discussed which of these parameters are suited for personalisation, and how this can be done. The proposed approach is then presented with its pros and cons, which results in a list of contributions this project has to the field of personalisation.

1.1 THE EDITORIAL MIX PROBLEM

In the conventional newspapers the editors job is to compose an intriguing front page that offers the contents of the sections that might interest the individual user. His challenge is to accommodate the needs of the newspapers segment of readers, divide the articles into sections, with a nice reading flow and attractive illustrations, and hand-pick articles to go on the front page. In 1965 [Haskins, 1965] defines the editorial mix problem as finding the least number of items to obtain maximum audience coverage, but with adaptive web sites it is possible to obtain a single user's preferences and accommodate them. Therefore it is possible to redefine the problem to:

Finding the composition of articles that provides the best satisfaction of the individual user preferences.

An important part of the editorial mix is also that each piece of the mix needs to be interesting in it self, as stated by [Tidwell, 2010] in her definition of the editorial mix pattern.

[Perkowitz and Etzioni, 2000] decomposes the problem of synthesising an adapted page into several subproblems¹ :

- What is the content (that is, set of items) of the index page?

¹ In the subproblems stated here, "hyperlink" has been replaced by "item" to appreciate them more generally, rather than the original specific sense.

- Does it have a coherent topic? What should its title be?
- How are the items on the page ordered?
- How are the items labelled?
- Is the page consistent with the site's overall graphical style?
- Is it appropriate to add the page to the site? If so, where?

Some efforts have been made to digitally calculate similarities between articles and based on a current article suggest similar reading material or use collaborative filtering to suggest articles based on other users reading behaviour. Some papers propose a composition of articles from user picked RSS-feeds, which can, e.g. in the case of Google Reader, be divided into sections. This comes close to conventional newspapers, but there is no ordering of the flow of articles. The ordering, flow and choice of relevant articles is from here on referred to as using *relational* features for creating the editorial mix.

The solution for some digital newspapers are still to have an editor to create their coherent composed digital newspaper, like the New York Times or Wired Magazine. Flipboard, on the other hand, composes their editorial mix of articles from feeds and divides their pages into three or more rarely four or five articles¹ with excerpts and images, much like conventional newspapers front pages. How they choose their composition is kept a business secret, but it does seem to vary a lot, see Figure 1.

¹ Flipboard includes specialised layouts with more articles per page for Twitter.

Maybe put in a reference to a conventional newspaper.

It is hard to say if there is a control behind the placement of content other than the choice of featured and non-featured articles, but this is actually an example of a computationally composed newspaper. The placement and amount of room given for an article is from here on referred to as using *spacial* features for creating the editorial mix.

Finally, subjects of articles have more relevance at some points in time than others and editors choose the amount of time stories should be available in, where RSS-readers just displays the



Figure 1: A screenshot of composition of three articles in Flipboard, with different subjects, i.e. world crime, world finance and technology news.

newest articles first, which are not always the most relevant. This selection of articles within a chosen time frame is from here on referred to as using *temporal* features for creating the editorial mix.

One thing that is vastly different from newspapers to RSS-readers is the ability to deliver personalised content, in that a user can choose which RSS-feeds to follow, whereas readers of newspapers need to navigate it in order to find interesting articles. Also where newspapers have quality assurance of its content, RSS-readers have a seemingly unlimited amount of articles.

1.2 PERSONALISING A DIGITAL NEWSPAPER

[Bush, 1945] describes a collective memory library machine that can be indexed, called the memex. Items in the library are linked together forming personal association trails. This is the early conception of the hypertext media that would later become the World Wide Web and later personalised web applications. With many respects that is what this project tries to achieve; i.e. link information in the form of articles together

and present them in personalised trials defined by the user. As opposed to [Bush, 1945] proposed manual linking it is now possible to, e.g. classify and compute similarity automatically, which greatly aids the process.

User preferences are very diverse and it is therefore hard to accommodate every individual in a single solution. A digital solution must be bound to a specific domain, but must also be open for novel use.

"Web personalization is defined as any action that adapts the information or services provided by a Web site to the needs of a particular user or a set of users, taking advantage of the knowledge gained from the users' navigational behavior and individual interests, in combination with the content and the structure of the Web site."

– [Eirinaki and Vazirgiannis, 2003]

The three categories of the editorial mix can be described in the sense of personalisation as well. Accommodating individual user preferences based on spacial personalisation is achieved by a placement of articles, temporal personalisation by selecting articles of higher news value based on their relevance time frame and finally, relational personalisation by selecting articles that provides more value based on their respective and collaborate topics. Temporal personalisation is also obtained by letting user preferences have a life time and decrease the preference influence on which articles to select as time passes. This is referred to as personalising using a temporal user model, whereas the former is referred to as personalising temporal user preferences.

All three categories are related as they each provide some value to the editorial mix; a different spacial placement of a specific article can, e.g. provide a different composition of the editorial mix and therefore a different relational value to the user, which also means that the user will discover articles at different times and therefore also provide different temporal value.

This paper seeks a more general approach to solving personalisation problems, and tries to establish the contributions of Constraint Programming to this field.

1.3 CONTRIBUTIONS OF CONSTRAINT PROGRAMMING

As a declarative programming language, Constraint Programming (CP) offers means for describing the problem to be solved using constraints and a general purpose constraint solver. Once the general purpose solver is set up, the constraints can be defined to model the problem to be solved, but does not necessarily make it easy. However, the problem definition can easily be extended and modified afterwards.

"Ordinary people generally aren't interested (and rightly so) in low-level programming details – they just want to express the problem in some reasonably congenial way and let the system get on with solving the problem. [...] Having to deal only (or mostly) with the logic component simplifies many things for the programmer. First, (the logic component of) a declarative program is generally easier to write and to understand than a corresponding imperative program. Second, a declarative program is also easier to reason about and to transform, as much current research in functional and logic programming shows."

– [Lloyd, 1994]

The editorial mix, and personalisation problems in general, consists of a series of requirements on what should be shown to the user. These requirements changes according to the individual user, often just by an adjustment of a general requirement. Because CP is a language for modelling requirements, it can be a great contribution to this field. If requirements are modelled as logic constraints, the changes to the individual user could be done by adjusting variables and parameters of the given values to fit the individual user.

To be able to work with personalisation problems in CP, the problems of which CP works on needs to be defined. Constraint Satisfaction Problems (CSPs) and Constraint Optimisation Problems (COPs) are the two types of problems CP can be used to solve. The following descriptions of CSPs and COPs have been modified to fit personalisation problems from the original definitions provided by [Russell and Norvig, 2003] and [Apt, 2006].

A Constraint Satisfaction Problem is defined by the 4-tuple $(\mathcal{V}, \mathcal{X}, \mathcal{D}, \mathcal{C})$, where \mathcal{V} is the set of values, \mathcal{X} is the set of variables, \mathcal{D} is the corresponding set of domains and \mathcal{C} is the set of constraints on the variables.

Each variable has a corresponding domain and each domain has sub-domains corresponding to the sub-variable, or attribute, of each variable. A value has a set of attributes, that may extend that of a variable, but if a variable is assigned, the variable's attributes should reflect that of its assigned value.¹ Unlike [Russell and Norvig, 2003] and [Apt, 2006] a constraint is here defined as a function on specific variables returning a boolean value. Therefore the tuple of u values with w attributes, n variables with m attributes and p constraints, where the i th constraint is defined on s_i variables, can be expanded to:

$$\left(\begin{array}{l} \mathcal{V} : \left\{ v_1 : \begin{pmatrix} v_{1.a_1} \\ \vdots \\ v_{1.a_w} \end{pmatrix}, \dots, v_u : \begin{pmatrix} v_{u.a_1} \\ \vdots \\ v_{u.a_w} \end{pmatrix} \right\}, \\ \mathcal{X} : \left\{ x_1 : \begin{pmatrix} x_{1.a_1} \\ \vdots \\ x_{1.a_m} \end{pmatrix}, \dots, x_n : \begin{pmatrix} x_{n.a_1} \\ \vdots \\ x_{n.a_m} \end{pmatrix} \right\}, \\ \mathcal{D} : \left\{ d_1 : \begin{pmatrix} d_{1.a_1} \\ \vdots \\ d_{1.a_m} \end{pmatrix}, \dots, d_n : \begin{pmatrix} d_{n.a_1} \\ \vdots \\ d_{n.a_m} \end{pmatrix} \right\}, \\ \mathcal{C} : \left\{ c_1 : func(x_{(1,1)}, \dots, x_{(1,s_1)}) \rightarrow \mathbb{B}, \right. \\ \quad \quad \quad \vdots \\ \left. c_p : func(x_{(p,1)}, \dots, x_{(p,s_p)}) \rightarrow \mathbb{B} \right\} \end{array} \right) \quad (1)$$

Where \mathbb{B} is either true or false.

A CSP is a special case of a Constraint Optimisation Problem (COP) and a COP is defined by the 5-tuple $(\mathcal{V}, \mathcal{X}, \mathcal{D}, \mathcal{C}, \mathcal{O})$, where the three first elements are defined as in a CSP and \mathcal{O} is a set objective (or cost) functions on variables, that determines the quality of a current state. The set of objective functions can be described with the same structure as constraints in CSPs

¹ Others do not consider representation of values because they in their case only consist of simple integer, real or boolean values.

and can be expanded as follows with q constraints, where the i th function is defined on t_i variables.

$$\mathcal{O} : \left\{ \begin{array}{l} o_1 : func(x_{(1,1)}, \dots, x_{(1,t_1)}) \rightarrow \mathbb{R}, \\ \vdots \\ o_q : func(x_{(q,1)}, \dots, x_{(q,t_q)}) \rightarrow \mathbb{R} \end{array} \right\}$$

Where \mathbb{R} is the set of real numbers.

Satisfaction (or regular) constraints are also called hard constraints and objective functions are called soft constraints because a solution can be found if all hard constraints are satisfied, whereas an optimal assignment is enough to satisfy objective functions.

Finally, a constraint on a single variable is called an unary constraint, a constraint on two variables is called a binary constraint and a constraint on three or more variables is called a global constraint.

1.4 PROBLEM DESCRIPTION

Tilføj editorial mix til de første linjer af problemformulering så det er skåret ud i pap at problemet vel primært er hvordan man kan personalisere the editorial mix og dernæst hvorvidt constraint programming kunne være en løsning.

This project is a feasibility study of the implementation of CP in the field of personalisation. The criteria of success is whether it is possible to successfully implement general techniques of personalisation using CP and to make personalisation more accessible with the aid of CP. Therefore the project will be divided into two main areas; i.e. an assessment of the use of CP in the context of personalisation and a direct application of this in the form of a personal digital newspaper, where CP is used to personalise the content and composition of a digital newspaper.

Many techniques for personalising digital solutions already exists, but the role of CP within this domain has not been determined. This project seeks to explore CP as a tool to make the personalisation of digital solutions.

1.4.1 *Personalisation Challenges*

In an attempt to introduce a personal editorial mix in the digital newspaper, the report will try to analyse which preferences the users will have with respect to the content, composition and the time frame of relevant articles. It will describe the search for articles to fit the user needs as an Constraint Optimisation Problem and try to solve it. What makes a newspaper is not only the accumulated content of its articles, but the arrangement of them. “Which articles should go where” is just as important, and the placement of articles in the newspaper should therefore go through an equal solving process.

1.4.2 *Algorithmic Challenges*

To be able to use and assess CP in the context of personalisation a full understanding must be acquired. Features that can be solved using CP will be modelled as a COP and solved. Furthermore, because the problem has a fixed budget for finding a solution, its algorithmic complexity will be analysed. The findings will be concluded in an evaluation of the applicability of CP to personalisation problems.

2 | PERSPECTIVES AND INSPIRATIONS

Web personalisation is by [Mobasher, 2007] divided into phases of data collection and preprocessing, pattern discovery and evaluation, and applying the discovered knowledge in real-time to mediate between the user and the Web. There have been many suggestions on how to tackle these different processes of creating the interactive personalised digital newspaper. [de Buenaga Rodríguez *et al.*, 2004] proposes a strictly stochastic approach to dynamic personalisation obtained by characterisation of content and user's interests. Both implicit and explicit relevance feedback¹ is used to refine the user models. Stochastic approaches have the advantage of being effective, but often solves a very specific problem. Also, these approaches tend to get very complex in order to deliver promising results. Some cope with this by introducing logic to the problem like it is done in [Nilsson, 1999] with a spacial approach. In this project it is possible to benefit from the structure of the logic approach of CP and the effectiveness of a stochastic approach by introducing preference constraints with an objective function.

Reference for stochastic approaches?

Many uses the approach of computing the tf-idf similarity with a cosine function as the distance function, as it is done in [Díaz and Gervs, 2005] to apply relational personalisation. It is based on a set of keywords extracted from the news items and a set of training documents. This constitutes the initial approach for computing similarity in this project. However, [Abuzir and Vandamme, 2002] argues that semantic knowledge is more substantial than keywords.

Classification techniques can be applied in order to ease the task of selecting relevant articles and determine their mutual

¹ Implicit is when the (unaware) user's behaviour is recorded to determine relevance and explicit is where the user is aware of the action of giving the feedback.

relationships. [de Buenaga Rodríguez *et al.*, 2004] uses a library of documents to train a categorisation algorithm and the users are then asked to select categories of which they have interest. [Abuzir and Vandamme, 2002], on the other hand uses a thesaurus of hierarchically, and to the task specifically, structured terms to index news articles. Results of the indexing are thereafter mapped with user profiles to select the relevant articles. In stead of using predefined root terms as the basis for a classification, WordNet can be used to obtain semantic knowledge for a document. WordNet is a large lexical database of English words and their relationships in the form of different graphs. [Bouras and Tsogkas, 2010] presents an algorithm for enriching articles using WordNet's hypernym-graphs. WordNet also contains similarity functions between words. These functions will later on constitute the next step for computing similarity in this project.

[Díaz and Gervs, 2005] does present the means of combining the use of categories and keywords, but this approach demands predefined categories, which must be kept updated in order to follow semantic changes to the field. The time limitations and prioritisation of this project did not allow for a thesaurus to be obtained to aid the classification and will therefore not be introduced to the solution. One, could also argue that semantic assumptions are made, when categories are predefined, which could lead to some false classification. Whereas the structure of [Bouras and Tsogkas, 2010]'s algorithm bases its semantic structure only on words from the article and the general semantic (and more neutral) structure that constitutes the basis for WordNet.

[Claypool *et al.*, 1999] presents a combination of content-based and collaborate filters to predict interest in articles based on a user profile.

[Claypool *et al.*, 1999] presents a front page design and available sections. A section can appear as their front page. relevance feedback

their editorial mix consists of ordering the articles by most predicted interest first.

[Díaz and Gervs, 2005] along with both a short- and long-term representation of the user models. Furthermore, a global user profile, to get the process of generating the user model started.

[Centeno *et al.*, 1999] incorporates temporal personalisation in that it is possible to ask for articles in the newspaper based on a specific period, but also by incorporate ageing of user interests.

[Esteban *et al.*, 2001] incorporates temporal features of their personalisation of a digital newspaper based on Yahoo! Spain. Automatic Categorisation of news items, long- and short-term user models.

In the explored literature users shows much interest in being able to turn pages as it is done in a regular newspaper. [Ihlström *et al.*, 2004] describes this as “open, turn pages, chose article, read and return”.

[Díaz and Gervs, 2005] proposes the use of collaborate filtering to handle the problem of converging, which is what will happen if no non-personalised articles are introduced, but this still only concerns articles that are within the area of the users interest. If e.g. a user has not shown interest in politics, the news of Barack Obama becoming the President of USA will never be included in the newspaper. Instead a ratio between personalised and general articles will solve this issue, and since it is not within everyones interest to receive general news, this ratio should be adjustable.

Users navigate the newspaper using sections and headlines as the main entry points [Ihlström *et al.*, 2004] and these should therefore be kept in the digital version.

Users express that these should be put into menu [Ovesson and Wikström, 2005].

[Díaz and Gervs, 2005] proposes personalised excerpts from the articles to further ease the navigation.

There exist many different examples of preference modelling using CP, and [Abidi and Chong, 2004] describes a factual information system to find personal information, e.g. about healthcare. They present two constraints: (1) select only

information-objects that correspond to the user-model and; (2) the content of the retained information-items do not contradict each other. This is an example of an editorial mix in that it incorporates the relational features, i.e. both between user and articles, and articles in between. However, they do not take into account the spacial part of their editorial mix, nor do they take into account any temporal features of the information needed. It is of cause notable that the user needs for a strictly factual information system are different than from a newspaper.

Another application of preference modelling using CP is [Vossen, 2005]. He proposes a CP approach to automatic playlist generation, which very much relates to what this project attempts to achieve. A playlist can be perceived as a, in this case, personal mix of songs. He presents constraints to exclude songs with certain attributes, to model that certain songs should be similar to each other or a user preference, to describe preference about the number of songs from a specific artist and the well-known all-diff constraint. These can be directly translated to the editorial mix of a newspaper, where the songs are articles and artists could be a specific author or content provider.

som afslutning på kapitel 2 kunne du opsummere hvilke features du mener er mest relevante så det ikke bare bliver en lang liste

3 | FEATURES TO BE PERSONALISED

This section analyses the user preferences and identifies which features should be modelled as personalisation constraints. It also seeks to define the default setting that consists the general model for a good editorial mix of a digital newspaper.

The focus of automatically generating the editorial mix introduces temporal, spacial and relational circumstances about the composition. But before these can be modelled as constraints it is necessary to look at the user needs of the application.

Some prerequisites must be stated in order to set the focus of the problem, but also for the potential user to relate to the product. The research done by [Ihlström *et al.*, 2004] and [Ovesson and Wikström, 2005] determines the preferable size of the digital newspaper to be $14.732 \times 20.828\text{cm} \sim \text{size A5}$, which reflects the size of the iPad. Therefore this project will target the iPad as its primary device. The hand-held device also introduces mobility, which is also of great preference to the potential users.

3.1 USER NEEDS

This section will define the user needs for the application. A full description of personas, scenarios and business case is found in appendix A [on page 55](#).

After the definition of the initial requirements was done the first prototype was developed. Its main features followed the requirements on turning pages, choosing an article, read it and returning to the overview of articles, see Figure 2a.

After a preliminary small survey it became clear, that a column-based layout would be more attractive, and would provide a better opportunity to explore the editorial mix. With a column-

Menu Front page Section 1 Section 2 Section 3 Search Columns

Front Page This section is about...

BBC News - Home

David Cameron: "What happened is completely unacceptable"

David Cameron has declined to reveal whom he has invited to dinner at his home - after a request made in the wake of the donor row - because such details are "private", No10 has said. But the PM promised a "proper inquiry" into donations after the resignation of Tory co-treasurer Peter Cruddas.

Mr Cruddas quit after reporters filming him laying donations to the Tory party were given access to Mr Cameron.

The BBC understands guests to the PM's home have included some party donors.

But ministerial aides at Number 10 say it will not reveal any further details about the guest list because the flat above 10 Downing Street is used as his private family home and Mr Cameron does not pay the £10,000-a-month rent there.

The known guests include long-standing party donors such as former Tory treasurer Michael Spencer and the party's current chief executive, Andrew Feldman, who is an old friend of Mr Cameron and his family.

Labour is conducting an independent inquiry after Mr Cruddas's claims, which were filmed by undercover Sunday Times reporters, came to light over the weekend. The matter has also been reported to the Metropolitan Police.

Labour had already accepted a disclosure of dinner guests announced by Mr Cameron on Downing Street.

And Tony MP Mark Field, a list of diners who had been entertained in the flat at No10 should be released.

Mr Cameron has pledged to hold a party inquiry into the claims, which he described as "completely unacceptable".

But Labour leader Ed Miliband said that was not good

InfoWorld News

March 26, 2012

Follow @gjdrill

The Eclipse Foundation for open source development tools is eying up a July 1 release date for the 1.0 version of its Orion browser-based IDE for building Web applications, which will be discussed at this year's EclipseCon 2012 conference in Rostock, Va.

Unlike the signature Eclipse desktop IDE, which is targeted to Java and C++, Orion is intended for Web application development, said Ian Skerrett, Eclipse vice president of marketing, in an interview. "The benefit [of Orion] is your development environment is where your code is running," he said. Orion is particularly useful for cloud application development, Skerrett added.

1.0 is the first major release that puts JavaScript on the server, as well as be featured at EclipseCon in a session about the *Vlet* Java-based IDE, which is in Eclipse's *Developer World* newsletter for more perspectives on software development.

BBC News - Home

David Cameron: "What happened is completely unacceptable"

David Cameron has declined to reveal whom he has invited to dinner at his home - after a request made in the wake of the donor row - because such details are "private", No10 has said. But the PM promised a "proper inquiry" into donations after the resignation of Tory co-treasurer Peter Cruddas.

Mr Cruddas quit after reporters filmed him laying donations to the Tory party were given access to Mr Cameron.

The BBC understands guests to the PM's home have included some party donors.

But ministerial aides at Number 10 say it will not reveal any further details about the guest list because the flat above 10 Downing Street is used as his private family home and Mr Cameron does not pay the £10,000-a-month rent there.

The known guests include long-standing party donors such as former Tory treasurer Michael Spencer and the party's current chief executive, Andrew Feldman, who is an old friend of Mr Cameron and his family.

Labour is conducting an independent inquiry after Mr Cruddas's claims, which were filmed by undercover Sunday Times reporters, came to light over the weekend. The matter has also been reported to the Metropolitan Police.

Labour had already accepted a disclosure of dinner guests announced by Mr Cameron on Downing Street.

And Tony MP Mark Field, a list of diners who had been entertained in the flat at No10 should be released.

Mr Cameron has pledged to hold a party inquiry into the claims, which he described as "completely unacceptable".

But Labour leader Ed Miliband said that was not good

InfoWorld News

March 26, 2012

Follow @gjdrill

The Eclipse Foundation for open source development tools is eying up a July 1 release date for the 1.0 version of its Orion browser-based IDE for building Web applications, which will be discussed at this year's EclipseCon 2012 conference in Rostock, Va.

Unlike the signature Eclipse desktop IDE, which is targeted to Java and C++, Orion is intended for Web application development, said Ian Skerrett, Eclipse vice president of marketing, in an interview. "The benefit [of Orion] is your development environment is where your code is running," he said. Orion is particularly useful for cloud application development, Skerrett added.

1.0 is the first major release that puts JavaScript on the server, as well as be featured at EclipseCon in a session about the *Vlet* Java-based IDE, which is in Eclipse's *Developer World* newsletter for more perspectives on software development.

DPR news

The prospect of gay marriage in the Evangelical Lutheran Church has led two bishops to propose completely renounce legally binding weddings from the church.

Colourbox

26 Mar, 2012 13:53 English

importance of the government's plan to allow same-sex couples to be married in the Evangelical Lutheran Church in Denmark, a number of bishops are now proposing completely removing legally

Front Page This section is about...

BBC News - Home

David Cameron: "What happened is completely unacceptable"

David Cameron has declined to reveal whom he has invited to dinner at his home - after a request made in the wake of the donor row - because such details are "private", No10 has said. But the PM promised a "proper inquiry" into donations after the resignation of Tory co-treasurer Peter Cruddas.

Mr Cruddas quit after reporters filmed him laying donations to the Tory party were given access to Mr Cameron.

The BBC understands guests to the PM's home have included some party donors.

But ministerial aides at Number 10 say it will not reveal any further details about the guest list because the flat above 10 Downing Street is used as his private family home and Mr Cameron does not pay the £10,000-a-month rent there.

The known guests include long-standing party donors such as former Tory treasurer Michael Spencer and the party's current chief executive, Andrew Feldman, who is an old friend of Mr Cameron and his family.

Labour is conducting an independent inquiry after Mr Cruddas's claims, which were filmed by undercover Sunday Times reporters, came to light over the weekend. The matter has also been reported to the Metropolitan Police.

Labour had already accepted a disclosure of dinner guests announced by Mr Cameron on Downing Street.

And Tony MP Mark Field, a list of diners who had been entertained in the flat at No10 should be released.

Mr Cameron has pledged to hold a party inquiry into the claims, which he described as "completely unacceptable".

But Labour leader Ed Miliband said that was not good

InfoWorld News

March 26, 2012

Follow @gjdrill

The Eclipse Foundation for open source development tools is eying up a July 1 release date for the 1.0 version of its Orion browser-based IDE for building Web applications, which will be discussed at this year's EclipseCon 2012 conference in Rostock, Va.

Unlike the signature Eclipse desktop IDE, which is targeted to Java and C++, Orion is intended for HTML and JavaScript development, said Ian Skerrett, Eclipse vice president of marketing, in an interview. "The benefit [of Orion] is your development environment is where your code is running," he said. Orion is particularly useful for cloud application development, Skerrett added.

1.0 is the first major release that puts JavaScript on the server, as well as be featured at EclipseCon in a session about the *Vlet* Java-based IDE, which is in Eclipse's *Developer World* newsletter for more perspectives on software development.

DPR news

The prospect of gay marriage in the Evangelical Lutheran Church has led two bishops to propose completely renounce legally binding weddings from the church.

Colourbox

26 Mar, 2012 13:53 English

importance of the government's plan to allow same-sex couples to be married in the Evangelical Lutheran Church in Denmark, a number of bishops are now proposing completely removing legally

Section 2 This section is about...

BBC News - Home

David Cameron: "What happened is completely unacceptable"

David Cameron has declined to reveal whom he has invited to dinner at his home - after a request made in the wake of the donor row - because such details are "private", No10 has said. But the PM promised a "proper inquiry" into donations after the resignation of Tory co-treasurer Peter Cruddas.

Mr Cruddas quit after reporters filmed him laying donations to the Tory party were given access to Mr Cameron.

The BBC understands guests to the PM's home have included some party donors.

But ministerial aides at Number 10 say it will not reveal any further details about the guest list because the flat above 10 Downing Street is used as his private family home and Mr Cameron does not pay the £10,000-a-month rent there.

The known guests include long-standing party donors such as former Tory treasurer Michael Spencer and the party's current chief executive, Andrew Feldman, who is an old friend of Mr Cameron and his family.

Labour is conducting an independent inquiry after Mr Cruddas's claims, which were filmed by undercover Sunday Times reporters, came to light over the weekend. The matter has also been reported to the Metropolitan Police.

Labour had already accepted a disclosure of dinner guests announced by Mr Cameron on Downing Street.

And Tony MP Mark Field, a list of diners who had been entertained in the flat at No10 should be released.

Mr Cameron has pledged to hold a party inquiry into the claims, which he described as "completely unacceptable".

But Labour leader Ed Miliband said that was not good

InfoWorld News

March 26, 2012

Follow @gjdrill

The Eclipse Foundation for open source development tools is eying up a July 1 release date for the 1.0 version of its Orion browser-based IDE for building Web applications, which will be discussed at this year's EclipseCon 2012 conference in Rostock, Va.

Unlike the signature Eclipse desktop IDE, which is targeted to Java and C++, Orion is intended for HTML and JavaScript development, said Ian Skerrett, Eclipse vice president of marketing, in an interview. "The benefit [of Orion] is your development environment is where your code is running," he said. Orion is particularly useful for cloud application development, Skerrett added.

1.0 is the first major release that puts JavaScript on the server, as well as be featured at EclipseCon in a session about the *Vlet* Java-based IDE, which is in Eclipse's *Developer World* newsletter for more perspectives on software development.

(a) Initial prototype layout with adjustable ratios between articles and a paged interface of each section.

Front Page

This section is about...

(c) Third iteration of the prototype with a column-based and “endless” layout. Sections are placed beneath each other.

www.washingtonpost.com

Friday, May 25, 2012

'Star Wars' Turns 35, Reminds Us How Much It Made Us Need Action Figures

Jen Chaney

On this day — May 25, 1977, the Wednesday before the Memorial Day weekend — “*Star Wars*” opened in theaters and changed the pop cultural landscape.

Our three “*Star Wars*” heroes. (Lucasfilm Ltd. & TM. All Rights Reserved)

To borrow the words of a Washington, D.C., resident whose Cleveland Park neighborhood was overrun that summer with “*Star Wars*” fans, we’re in the over and out to see Luke, Han and Leia at the Uptown Theater: “It’s ... it’s an invasion.” (Read this [whole Washington Post article](#) about the neighborhood consternation back in ’77. The quotes in it are a riot: “I told my wife, ‘Hey, some ... claim it’s ...’” followed by the driveway. “The funny thing is that it turned out to be a friend of my wife whose car was blocking the drive, a person who had just graduated from clown school.”)

wanted anything they could find that would allow them to re-create its universe of Stormtroopers, Death Star blow-ups and awesome singles bars on Tatooine. Thus, the action figure was born. And the radio-controlled R2D2. And a Death Star space station with a working trash compactor. Seriously, look at all this stuff Kenner successfully sold us.

There were book-and-record sets for kids. We wanted to read and hear the “*Star Wars*” story again, between their 178th and 179th visits to the theater to see it. There were many fast food tie-ins. (Man, I miss Burger Chef.) Also, I had no idea that, per this commercial, Alison Brie’s mother once worked there.)

Inevitably, the sequels were launched, which provided us with even more junk to buy. Like “*Empire Strikes Back* Underoos.” And breakfast cereal. And video games. And more toys, toys, toys:

(d) Third iteration of the prototype with a column-based and “endless” layout. Sections are placed beneath each other.

Figure 2: The figure shows three iterations of the prototype layout.

[August 2012 – Technical University of Denmark]

based layout the digital newspaper would have more resemblance to conventional newspapers and therefore it was possible to apply some of the same principles of the editorial mix.

After the second prototype was developed it was tested. A specification of the test can be found in Table 1.

Table 1: Test Specification

Test subjects	The test was conducted on a total of 7 test subjects of ages between 21-29, and of different sex and occupation.
Participants	Each test was done with 1 test conductor and 1 test subject.
Materials	An iPad with the application running and a computer to write notes on the test subject's statements and propositions.
Description	The test subject was presented with the prototype layout seen in Figure 2c and 2d. The test was conducted as an informal qualitative talk with a basis in the test subject's interests in such a product. Transcripts from each test can be found at http://lestrade.imm.dtu.dk/~s062596/data/test-transcripts.zip and a summary of the results in section A.5 on page 65.

The main points from the user test was that the newspaper should provide an overview of its contents, that it should be easy to navigate relevant new and archived articles, that the users wanted to provide relevance feedback on articles and indication of the relevance of the article. Moreover, the newspaper should provide a good balance between imagery and textual content, that white space in between articles is not a problem and finally that an article should be read screen by screen, even if it means dividing text into a new set of columns. One user in particular expressed that it had solved the problems that <http://nyhederne.tv2.dk/>¹ has and given the additional features it would provide a readable layout, easy navigation and a good overview of its content.

¹ The website of a Danish news channel.

The prototype consisted of the basic navigation between topic categories, i.e. sections, and articles. Navigational choices was made in order to present the general idea of the framework, but where more crucial choices on its uses have not been made yet. This was also to encourage the test subjects to talk about what uses they would have of the presented framework. How-

ever, they were also asked about the navigational structure and indeed some changes had to be done.

The above presented preparatory work resulted in the following user needs:

- Read articles presented in a nice and digestible layout
- Get an easy overview of the content of the newspaper
- Easily navigate between articles with few touch-friendly interactions
- Read relevant articles based on user defined topics

3.2 USE CASES

These user needs and the discussed requirements proposes some use cases from the application. Theses will presented here.

In order to fulfil the user needs, more general use cases are derived. These and the web servers handling of them are shown in Figure 3. The use cases shown in the figure is by no means a complete list, but they are selected because they have a closer relation to the user needs.

The user is able to select relevant topic categories to get an easy start with the application and his choices are thereafter saved to the user profile for later use. The user can select a section with a topic category from his selections and browse articles from this topic. If the user needs to get an overview of the content of the newspaper he can get the front page displayed, which contains a small collection of articles of most interest to the user. Moreover, it is possible to get a list of headlines contained within the current section.

In Figure ?? is shown a sequence diagram of what the system does in order to display the front page (or a section), when the user opens the application.

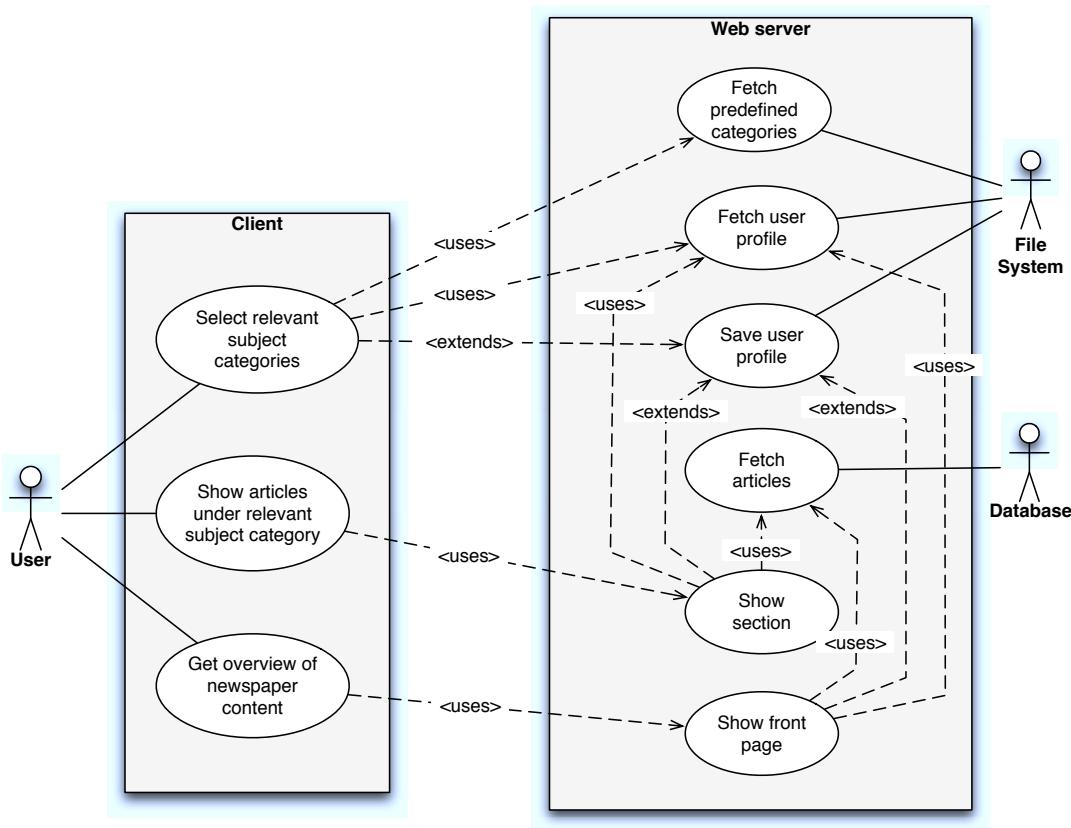


Figure 3: The figure shows the overall use cases of the system and how the web server acts to accomplish them.

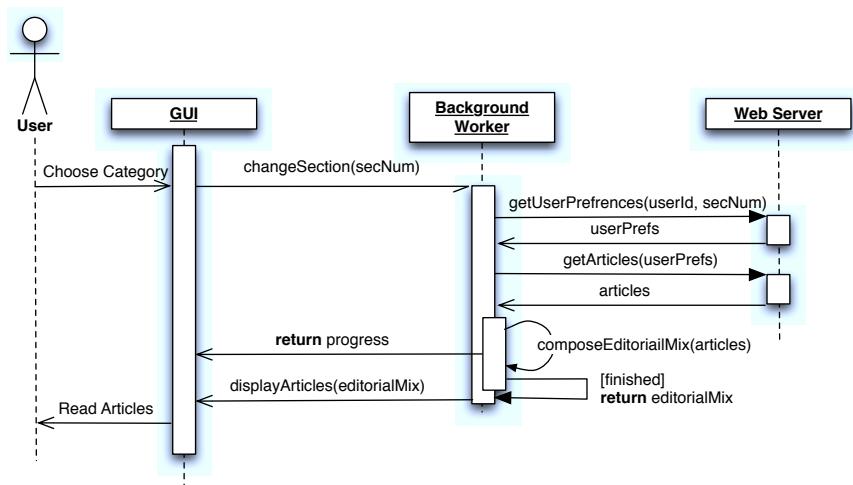


Figure 4: A sequence diagram from when the user chooses a topic category until reading articles. **secNum** is a section number, (front page is section 0), **userId** the user id, **userPrefs** the user preferences on a given section and **articles** is a library of articles to compose the editorial mix of.

When the application is opened, or the application is changed to display a section, a background worker is initialised to compose its containing mix of articles. Of course, if the mix of articles in a section, or the front page, have already been computed, it should not have to recompute it. The background worker needs to get both the user preferences of the chosen topic category and articles that potentially fit the user preferences. While the worker computes the editorial mix it sends messages to the user interface about the progress, to provide feedback to the user. When it finishes the user interface is asked to display the articles.

3.3 REQUIREMENTS

In the explored literature and the conducted tests users expressed some non-functional requirements. These are described in this section.

User needs states the requirement of having a clear overview of the content and as stated in [[Ihlström et al., 2004](#)], this includes a clear marking of the beginning and the end of the articles and sections. This is obtained by both having a summary of the most interesting articles on the front page and by having a list of headlines in each section.

From the user needs it is also required that the system should be easily navigated and as stated by [[Ovesson and Wikström, 2005](#)], this should be through click-able sections, headlines and through paging.

The layout, typography and design should be familiar to what is found in conventional newspapers, as stated by [[Ihlström et al., 2004](#)] and [[Åkesson et al., 2005](#)]. This is achieved by choosing a structure that resembles that of a newspaper and displaying content in balanced columns. Moreover, the user should be able to read the newspaper screen by screen, meaning that trailing text should be put into a new set of columns whenever it exceeds the screen, see Figure 5 and 6.

Maybe use “The newspaper indexing is the most effective ‘navigational’ tool in newspapers and headlines are the main entry points to text” [Ihlström *et al.*, 2004] instead?

In appendix A.4.2 on page 63 is found calculations on how many columns should be used on the iPad and on desktop computers based on conventional newspapers. The result of the calculations is that there should be 2 columns in portrait mode and 3 columns in landscape and on desktop computers of 1200px. However, it only requires a screen of 1320px before 4 columns would be optimal. But as the project targets the iPad screen size only 2 and 3 columns will be considered from here on.

The typography of the application should also resemble that of a conventional newspaper. It should contain a good ratio of both graphical and textual content and should, when possible, supply multimedia content. The conclusions made from the empirical data in [Ihlström *et al.*, 2004], about exploring which features to bring from conventional to digital newspapers, was that valuation and position of the news was important. More importantly that the reader should be guided through the digital newspaper. It is, however, crucial to consider that their empirical basis is not very large. They have a qualitative selection of respondents from newspapers that have, recent to its execution, become dedicated to the project. Moreover, they have chosen 16 open questions for the respondents to answer, which should provide some sort of basis for their conclusions. In this project it is chosen to use them as guidelines, but the choices made on this basis must be verified¹.

That the reader should be guided through the newspaper using valuation of the items does, however, fall in line with the editorial mix, which also have been used in conventional newspapers for a long time. This suggests control of the temporal, spacial and relational values of individual articles and between them. Also, using the Gestalt principles [Tidwell, 2010] suggests providing a visual hierarchy so the user can see the relative importance of the page elements and the relationship among them.

Finally, the respondents from [Ihlström *et al.*, 2004] also suggests that the paper should be continuously updated, but ac-

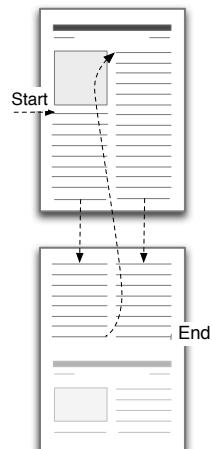


Figure 5: Reading pattern where the user has to scroll in order to see the full length of the column.

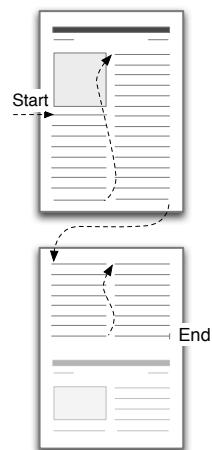


Figure 6: Reading pattern where the user can finish reading a whole page before scrolling to read the next.

¹ Earlier statements from this paper have been backed by additional sources.

cording to the user tests, it should be chosen how often this should happen.

Some technical requirements have also been gathered from the explored literature.

The navigation using headlines and article excerpts, suggests that these should be clearly marked and that also the excerpts should be personalised as suggested in [Díaz and Gervs, 2005]. The results from [Ovesson and Wikström, 2005] suggests that the menu with clickable sections should be placed on the left side of the screen, but they could have been biased as it was already placed there in the tested prototype. In, addition, they found this as a good choice as they recognised it from the web. A menu in the top of the page would therefore also be in line with their findings, as it is a general pattern of the web [Tidwell, 2010]. In addition, it would take up less space in the view, leaving more room for the general purpose of the application, namely reading. The menu items will work well as the user defines the content of them. Furthermore, it would aid the user to relate more to these divisions if it is possible for him to name them himself. Some problems may arise if there is not room enough in the top menu, e.g. when in portrait mode and, as the test subjects specifically requested, this could be solved by introducing a carousel-like arrow buttons to scroll the tabs, or even just using touch interactions.

Furthermore, it should be possible for the user to get an overview of the headlines contained in a section. This could be done by just having a list of the headlines, or by using the overview plus detail pattern presented by [Tidwell, 2010]. This could be done as in Sublime Text 2¹, see Figure 7

This overview could then show a larger (and readable) scale of headlines and the user should be able to see the images further down in the newspaper.

Many articles discuss different ways of represent the user's interests. It seems, however, that both [Díaz and Gervs, 2005] and [Billsus and Pazzani, 2000] generate good results with a dynamic short-term user model in combination with a static long-term user profile.

¹ Sublime Text 2 is a text editor for coding.



Figure 7: The figure shows the overview plus detail function in Sublime Text 2.

Rewrite to fit list. Maybe not divide into functional and non-functional.

Finally, the implementation of a community in the application should be done with the possibility of sharing the story on different social networks, but could also include comments on articles, as requested by the test subjects.

These requirements can be summed up in the following list:

- Summery of few most relevant articles on the front page
- Clear section and article headlines and personalised article excerpts to ease navigation
- Menu of section, that is always visible
- Paged navigation through sections
- Overview of article headlines
- Layout, typography and design should be familiar to a newspaper
- Visual hierarchy
- Balanced columns should be divided into screen size chunks
- Serve multimedia content
- The reader should be guided through the newspaper using valuation of the items in terms of categories of the editorial mix
- Combination of long-term and short term interest model of the user
- Incorporate community and social networking

3.4 THE EDITORIAL MIX

The task at hand is to decompose the spacial, temporal and relational features of the composition of articles that provides the best satisfaction of the individual user preferences into constraints. This section analyses the existing literature on reading behaviour of conventional and digital newspaper and derives constraints to compose the editorial mix of.

The reading behaviour of conventional newspapers differs from them read digitally. In the experiments done in [Holmqvist *et al.*, 2003] it is concluded that the net paper¹ readers read stories thematically close to their own specific profession or interests. So it is important to provide this setting for the reader. The readers also used the front page as a provider of main entry points. And finally, the readers "claim to scan more in order to find the two or three stories they will read in the net paper" which is explained by the poorer chances of links catching reader interest. However, it could be possible to attract the reading behaviour from conventional newspapers onto digital platforms - it is certainly interesting to see an equal analysis of the reading behaviour of tablet computers, which calls for more in-depth reading. If a digital platform are to attract more in-depth reading it requires some flow in the presentation of the articles, i.e. it requires an editorial mix, so the readers do not feel like they have left the main trail [Holmqvist *et al.*, 2003].

To attract reading behaviour of conventional newspaper it is worth while understanding readers expectations of these. [Holsanova *et al.*, 2006] confirms a summary of reading behaviour assumptions from [Kress and others, 1999] on conventional newspapers using eye-tracking measurements:

- Readers prefer the most general information at the top and the most specific information at the bottom of the semiotic space.
- Readers look for the most important information in the centre of the page and less important information on the periphery.
- Readers look for paratexts¹.

¹ [Genette, 1997] defines paratext as those productions accompanying a text, such as an author's name, a title, a preface, or illustrations.

And, two are not confirmed, but not declined either:

- Readers look for graphically salient elements; however, it is important to bear in mind that ‘what is made salient is culturally determined’ [Kress and others, 1999].
- Readers follow elements connected to each other by framing devices such as lines and arrows.

That the most important information should be in the centre of the page will be hard to attract on digital platforms because of the limited space. Because of the screen size only one or two, and in some cases three, articles are shown at a time and the user will have to scroll to see the next items. However, the relation between adjacent articles can still be controlled, so a featured¹ article should be adjacent to some smaller, but still very relevant articles. This will hopefully attract the same behaviour, but of course needs to be confirmed.

Also, a central position is hard to obtain, as many articles will be listed below each other, so a central position is here deemed to be higher than its relevant non-featured articles.

Based on the user study and the explored literature on reading behaviour the editorial mix problem can be divided in two; (1) the front page and (2) the sections.

¹ A featured article means providing it with more space than others, a central position and it is often accompanied with graphically salient elements.

1 The purpose of the front page is to draw attention and provide an intriguing overview of the whole newspaper. This is done by using many images and providing headlines and excerpts of the most relevant articles of the newspaper. The most relevant article should be featured in the centre with a selection of a little less, but still very relevant articles adjacent to it. A visual hierarchy should be provided so the user can see the relative importance of the page elements and the relationship among them. The front page should, if available, provide interesting articles from all sections as main entry points.

2 The purpose of each section is to provide a flow of articles relevant to a, by the user provided, topic that keeps the user interested and invites for in-depth reading. More general articles

should be placed in the top of the screen and more specific at the bottom, with a featured main article in the centre. Framing and lines should guide the user to what is related and a graphical .

These descriptions can be decomposed into the following constraints.

GENERAL CONSTRAINTS

- A featured articles should be allowed to take up more space
- A featured articles should be accompanied by an image
- A featured should have a central position
- A non-featured article should take up less space
- A featured article should be adjacent to non-featured articles
- All articles should be different

FRONT PAGE CONSTRAINTS

- Every article should have a very high level of relevance to at least one of the section topics
- Most or every non-featured article should be accompanied by an image

SECTION CONSTRAINTS

- Every article should have a high level of relevance to its containing section topic
- A section should contain an article if the front page contains the article and its relevance to this section is highest
- Articles should be grouped into subjects

- The section should contain a balanced weight between graphical and textual content
- Images should be spread evenly in the section

Part II

USING WHAT WE HAVE LEARNT

Text for part 2.

4 | PROPOSED DESIGN

This section describes the different choices in the interface for a digital newspaper and how the application can provide the user with an personal editorial mix.

Before it is possible express the problem formally the interface of the application must be described. This is done in the two following sections.

4.1 LAYOUT AND TYPOGRAPHY

When the user is at first presented with the application he should have as a direct path as possible leading to actually reading articles, which is anticipated that the user wants. Otherwise, there is no use for such an application. In this application is presented with a form to make choices about the contents of his personal digital newspaper. The application provides the possibility for choosing whether the front page should be visible or not. Moreover, the user can choose the topic for the first section from a list of predefined topics. If the topic he is looking for is not in the list he can choose to fill out some keywords and weights to cover his interests. After this he can provide it with a name for the section and choose to add another section or save his user profile. It is also possible for him to choose how many articles he would like in each section, including the front page. Figure ?? shows a mockup of the form.

modal, form (settings editor p.59 + grouping related items p. 134),

Figure 8: The figure shows a mockup of the form the user is presented when opening the application the first time.

**Figure 9:**

The figure shows mockups of the layout in landscape and portrait mode, respectively.

4.2 INTERACTIONS

To be able to solve the editorial mix problem using CP it must be expressed as a COP and the presented constraints must be translated to logical constraints. This is done in the following section.

4.3 PROBLEM REPRESENTATION

The division between the front page and the section can be kept in the problem representation, which will, as we will see later, be the source for the possibility of incorporating lazy loading of each. This section will therefore present a general problem specification for a section, which can be used in every section and on the front page, with varying constraints. Also, the front page will through this section and the following be referred to as a section, and specifically as section 0.

The set of values, \mathcal{V} , from equation 1 on page 8 in the problem is represented by a library of currently available articles and the set of variables, \mathcal{X} , is represented by the available positions in the section. Each variable can then be assigned a value in

the form of a specific article and when a complete assignment satisfies all constraints, a solution has been found. An article consists of a set of attributes, e.g. a date, the number of words in the article and a number indicating a relevance. Constraints are defined on variables and through them bound to their specific places in the section.

In the following the constraints presented in section [3.4 on page 26](#) will be formulated as logical constraints divided general constraints for all sections and specific constraints for the front page and other sections, respectively. Constraints defined on variables with letters a to d , means that the constraint is defined for every combination of variables from the problem. Hard constraints returns whether it is satisfied and preference constraints returns a violation, where 0 means not violated.

GENERAL UNARY CONSTRAINTS

$$\left\{ \begin{array}{l} \text{time-frame}(x_a) \rightarrow x_a.\text{date} \geq \text{today} - 7, \\ \text{featured-space}(x_a) \rightarrow x_a.\text{featured} = \text{false} \vee \\ \quad x_a.\text{columns} = 2 \vee \\ \quad x_a.\text{columns} = 3, \\ \text{nonfeatured-space}(x_a) \rightarrow x_a.\text{featured} = \text{true} \vee \\ \quad x_a.\text{columns} = 1 \vee \\ \quad x_a.\text{columns} = 2, \\ \text{featured-image}(x_a) \rightarrow x_a.\text{featured} = \text{false} \vee \\ \quad x_a.\text{has_image} = \text{false} \end{array} \right\} \quad (2)$$

Where n is the number of positions and therefore variables, in the section and today is variable that holds the current date. The `time-frame` constraint is set to include articles from a week ago, but can of cause be adjusted. As the layout is defined in 2 and 3 columns

GENERAL BINARY CONSTRAINTS

$$\left\{ \begin{array}{l} \text{featured-adj}(x_a, x_b) \rightarrow \text{not adjacent}(x_a, x_b) \vee \\ \quad (x_a.\text{featured} = \text{false} \wedge x_b.\text{featured} = \text{false}) \vee \\ \quad x_a.\text{featured} = \text{true} \wedge x_b.\text{featured} = \text{true}, \\ \text{featured-pos}(x_a, x_b) \rightarrow \text{not adjacent}(x_a, x_b) \vee \\ \quad (x_a.\text{featured} = \text{false} \wedge x_b.\text{featured} = \text{false}) \vee \\ \quad (x_a.\text{featured} = \text{true} \wedge x_a.\text{position} > x_b.\text{position}) \vee \\ \quad (x_b.\text{featured} = \text{true} \wedge x_b.\text{position} > x_a.\text{position}), \\ \text{adj-subj}(x_a, x_b) \rightarrow \text{not adjacent}(x_a, x_b) \vee \\ \quad 64 \cdot \text{relevance}(x_a, x_b)^2 \\ \quad -64 \cdot \text{relevance}(x_a, x_b) \\ \quad +15.36 \end{array} \right\} \quad (3)$$

GENERAL GLOBAL CONSTRAINTS

$$\left\{ \text{all-diff}(x_1, \dots, x_n) \rightarrow \bigwedge_{i=1, \dots, n} \bigwedge_{j=1, \dots, n} \text{not } x_i.\text{id} = x_j.\text{id} \right\} \quad (4)$$

FRONT PAGE CONSTRAINTS

$$\left\{ \begin{array}{l} \text{relevance}(x_a) \rightarrow \bigvee_{i=1, \dots, g} \text{relevance}(x_a, i) \geq 0.75, \\ \text{nonfeatured-image}(x_a) \rightarrow x_a.\text{has_image} = \text{false} \end{array} \right\} \quad (5)$$

Where g is number of sections and $\text{relevance}(\text{variable}, \text{section number})$ returns the relevance of the given variable in the given section number.

SECTION CONSTRAINTS

$$\left\{ \begin{array}{l} \text{relevance}(x_a) \rightarrow \text{relevance}(x_a, k) \geq 0.65, \\ \text{nonfeatured-image}(x_a) \rightarrow x_a.\text{has_image} = \text{false}, \\ \text{fp-article}(x_1, \dots, x_n) \rightarrow \bigvee_{i=1, \dots, n} \bigvee_{j=1, \dots, m} x_i.id = a_j.id, \\ \text{image}(x_a, x_b, x_c, x_d) \rightarrow \bigvee_{i=a, \dots, d} x_i.\text{has_image} = \text{true} \end{array} \right\} \quad (6)$$

Where k is the current section number, $a_1, \dots, a_m \in fp_articles$ is a list of articles contained in the front page and $\text{relevance}(\text{variable}, \text{variable})$ is a function returning the mutual relevance between the two given variables.

These constraints are more or less simply translated from the presented textual constraints into logical functions often containing most or every variable in the problem. It is however, in many cases profitable to choose a different representation of the problem using, e.g tree decomposition or reduce the constraints to binary constraints. Lets first express the constraints as they are formulated and work on them to form a better representation.

From the description of the editorial mix it is possible to model General requirements: Whitespace between articles should be minimised. This is not an actual user requirement, but it is included as it a fairly complex layout problem to solve and it shows some aspects of what it can do.

4.4 DELIMITATION

The proposed design in chapter [4 on page 31](#) should of cause account for the presented requirements, but there are some requirements that will not be implemented due to prioritisation and time limits.

The application presented in the next part will account for the discussed user model, but will only describe how to collect the necessary data and use it in the application, and not implement it.

The social aspects of the system is of cause an important part of it and in terms of the business case, social media are easy channels for awareness. [Tidwell, 2010] even states her editorial mix pattern as a social media pattern. Nonetheless, these will not be implemented in the presented application as it does not contribute with new knowledge to field. The articles obtained for this project contained both images and videos, but only images will be considered here. It is however easy to implement the support for videos as the same space allocation principles applies.

Also, the personalised summaries have already been very well explored in the paper [Díaz and Gervs, 2005] and this project will not try to compete with it.

The categories used to compose sections of is not a full list, but are some of the most recurring in popular news sites and are used just in order to proof the concept is possible. Thus, their definitions are not comprehensive either.

Finally, the full list of presented editorial mix constraints will be included in the proposed design, but only a selection of them will be implemented.

5 | TECHNICAL SPECIFICATION

This section describes a technical specification of a direct implementation of the design choices made in the previous chapter in terms of front end and back end tasks.

5.1 CONSTRAINT PERSONALISATION LIBRARY

5.1.1 *Data Structure*

5.1.2 *General Purpose Solver*

5.2 FRONT END

- Dynamic Page (created from form)

css: conditional styling

5.3 BACKEND WORKER.JS

- Lazy Loading Sections - Knowledge Passing

worker i tråd for sig selv, men samme proces

worker.js to create a background worker to perform the constraint programming.

Model-View-Control using backbone.js

Paging: single page web apps + manipulation the browser history https://developer.mozilla.org/en/DOM/Manipulating_the_browser_history

Preference ordering of hard constraints or division between preference constraints and hard constraints.

Assignment from library in stead of arbitrary assignment? The latter is a more hypothetical approach. Providing the library as a constraint, where each variable assignment must have a unique combination from one of the possibilities of the constraint. (sim,breaking,chars,date,sections?,columns):list The former introduces an implicit constraint in that the general purpose solver can only choose from the library, thus can only choose a combination of values that exists.

Ranges can be optimised in space by converting them to integer ranges. This can be done by setting min = 0 and max = (b-a)/gap.

Furthermore each subdomain should be able to be represented by a set of ranges and atomic values. Propagating through values causes many iterations and a whole range may be discarded by looking at its maximum and minimum value. However, if the range holds a potential valid value (solution to a variable) it can be divided into smaller ranges and their minimum and maximum values may be examined. This divide-and-conquer technique may continue until the search reaches atomic values (determined by the gap value of the range). If some atomic values and ranges seems to fulfil the constraints they should be returned. And the subdomain now consists of both ranges and atomic values.

Optimal/promising fixed budget computation

The library could take any combination of constraints and then organise them into conjunctions of disjunctions, with the constraints taking fewer values first.

In the implementation this is done by hand, so the program takes conjunctions of disjunctions of constraints organised with constraints that takes fewer values first. Constraint weighing could also help organising the disjunctions and furthermore

lead the search to concentrate on variables that is bound by these constraints. (p. 222 AIRussel).

Constraints should point to specific variables, this makes it somewhat rigid/ineffective because I have to write at global constraint that accounts for everything (ineffective in propagation – might also be a problem if it does not show progress in changing values, i.e. it is a hard constraints and not returning a cost of the set of values.) or divide it into smaller constraints separated by an ‘or’ (v). The latter is ineffective because there would should be a combination of constraints accounting for every situation, e.g. if the first variable is satisfying an unary constraint, the next say 3 variables (if the problem holds 4 variables) could satisfy three unary constraints, an unary and a binary (two combinations exists) or a constraint that takes three variables. This grows fast with the number of constraints.

Does it make sense that a continuous range cannot have specific values removed? Should it be possible for it be divided into subranges if the user decides to remove a range of values in between its domain of [min,max]?

Pool of workers to compute sections and send results to another worker to handle front page articles. Or, a lazy load approach where the front page is computed and then sections are computed with hard constraints that manages placement of articles within the given sections, e.g. if an article from the front page has the potential to be placed in only one section the constraint should state this, but it would demand cross-worker-constraints to handle if an article from the front page has the potential to be placed in more section (i.e. xor). color

Reducing constraints to binary constraints.

Implement visual difference between featured articles and non-featured articles.

5.4 SERVER FOR ACQUIRING AND MINING DATA FOR PERSONALISATION

5.4.1 *Spacial, temporal and relational personalisation*

5.4.2 *WordNet Enriching Articles*

Using key words

The initial approach involved computing the tf-idf similarity between documents and the user and the documents in between using the Python libraries for this [Bird *et al.*, 2009]. This approach works on a bow (bag-of-words) with key words and weights representing a single item. The weight is computed by the number of occurrences in the provided text and a cosine distance determines the similarity. Python also provides an interface for working with WordNet – a large lexical database of English words and their relationships in the form of different graphs. This opens the door to a more in-depth analysis of the obtained news items. [Bouras and Tsogkas, 2010] presents an algorithm for enriching articles using WordNet’s hypernym-graphs. A hypernym graph is generated by the top 20% frequent keywords of an article and weighted by:

$$W(d, f) = 2 \cdot \frac{1}{1 + e^{-0.125(d^3 \frac{f}{TW})}} - 0.5$$

Where d stands for the node’s depth in the graph (starting from root and moving downwards), f is the frequency of appearance of the node to the multiple graph paths and TW is the total number of words used to generate the hypernym graph.

In order to be able to work with hypernyms, the words must be converted to synsets. For each word there exists a synset for each use of the word, with the most frequently used first. Every synset is included at this point, but in a later stage this could be further focused by only using the top n . An analysis on how many percent of the words

5.4.3 *Meta Data from Open Calais*

5.4.4 *Computing Similarity*

- Storing Data

Part III

LEARNING WHAT WE HAVE USED

Text for part 3.

6 | DISCUSSION

7

EVALUATION OF THE SOLUTION

Evaluate using the in [Diaz *et al.*, 2001] and [Esteban *et al.*, 2000] presented method. Their categorisation based on few keywords (5 as the lowest) to represent a category resulted in poor evaluation, this gives a good motivation for including more keywords and using Wordnet to enrich the set of keywords.

It could be interesting to evaluate the precision and recall points based on: news items per section, news items per category, maximum number of news items per message required by the user, general relevance of the contents of a given day for a given user, etc. as [Diaz *et al.*, 2001] proposes.

[Diaz *et al.*, 2001] also raise the problem of precision in finding news within based on a single day. This can hopefully be handled by having the user specify in which period of time he wants news and maybe notify the user that solutions might be inaccurate if a limited period of time is chosen, or just limit the user to specify 24 hours as a minimal value.

7.0.5 Initial Test

NOT ONLY 5 test persons [Nielsen, 2012]. This has been discarded by many.

7.0.6 *Result*

7.1 TEST

7.1.1 *Layout*

classify items as first and second articles and use layout to distinguish them.

columns constraints, white space and odd placement of articles.

We follow a strict vertical structure, but there is a lot of work to be done with the horizontal structure

In order to come closer to understanding what newspapers does it could be interesting to analyse their component structure, e.g. using [Liu *et al.*, 2001] algorithm.

At an early stage the paging of a section was discarded, because a preliminary test (ask around) showed that users wanted to scroll down to see the full section. This was also necessary if a full-length articles were to be shown, but the reading behaviour analysis suggests that articles should be divided into chunks of subjects, which may be better to visualise using pages. This way the featured article could be shown in a longer length excerpt and stories on the same subject could surround it with only headlines, images and short excerpts shown. If the user then wants to read one of the articles in full length he can select it and the full article could be displayed, using the full size of the screen.

7.1.2 *Content*

The system can be used for automatic classification of articles. Of cause, then a sufficient list of categories and their definitions must be used. This can either be retrieved by the list of

Google News categories¹ and a Wordnet enriched list of key words from Google News list of suggested keywords² or by the root terms presented in [Abuzir and Vandamme, 2002]. These can later on be refined by information retrieved by the user behaviour in the system and manually removal of false negatives? However, also more advanced techniques of text classification could be used in later stages of the system, like one presented in [Esteban *et al.*, 2000].

Maybe find a better example of text classification.

Use of automatic generation of personal item summaries [Díaz and Gervs, 2005]

Use geotargeting to supply local news.

Use a thesaurus and predefined root terms as in [Abuzir and Vandamme, 2002] which improves classification; semantic knowledge is more general than keywords.

scattered ads [Ovesson and Wikström, 2005]

7.1.3 *Functionality*

Order a print copy of the newspaper

The development of the Internet from a distributor of information to a library of digital applications has deeply integrated the users in every step of an applications lifetime. It has become harder to distinguish between super users and developers, applications are branched and modified according to every need and authors can therefore no longer predict which use his application can be to another user – nor should he have to.

¹ <http://support.google.com/webmasters/bin/answer.py?hl=en&answer=42993>.

² <http://support.google.com/news/publisher/bin/answer.py?hl=en&answer=116037>.

7.1.4 *Improvements*

[Centeno *et al.*, 1999] suggest virtual communities, or individuals with common interests.

Weights on key words should be adjusted by a strength (or an uncertainty) of prediction as it is proposed in [Claypool *et al.*, 1999].

Count the number of sources have included an articles that are very similar to find the breaking factor.

How to handle that a user is not presented with an already read article? [Billsus and Pazzani, 2000] presents the nearest neighbour (NN) algorithm approach, using a tf-idf similarity, to determine whether the story is already known, i.e. the similarity to the NN read story is above a given threshold. This could be solved by keeping a library of read items (this can be done along with the tracking of which article is in focus) and then match new items against this banned list and down prioritise them if their similarity is too high. This could be done with the same polynomial function as used between articles.

“Distinguishing between short-term and long-term models has several desirable qualities in domains with temporal characteristics (Chiu and Webb, 1998).” [Billsus and Pazzani, 2000].

8 | CONCLUSION

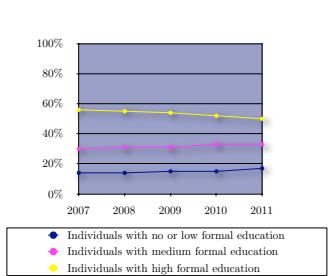
Part IV

APPENDIX

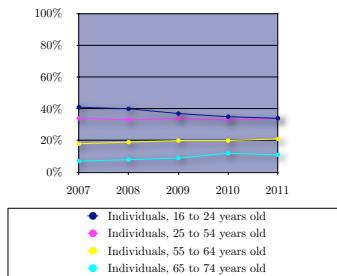
A | USER NEEDS

This section will define the user needs for the application.

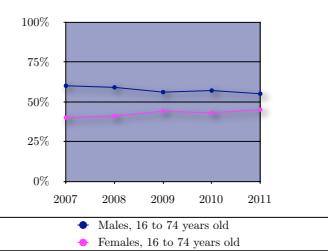
A.1 PERSONAS



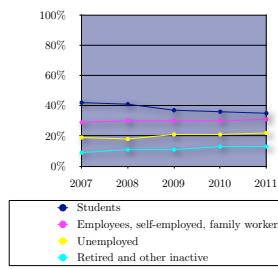
(a) Initial prototype layout with adjustable ratios between articles and a paged interface of each section.



(b) Second iteration of the prototype with an “endless” layout. Sections are placed beneath each other.



(c) Third iteration of the prototype with a column-based and “endless” layout. Sections are placed beneath each other.



(d) Third iteration of the prototype with a column-based and “endless” layout. Sections are placed beneath each other.

Figure 10: Eurostat: Individuals using the Internet for reading / downloading online newspapers / news magazines

In Figure 10 is shown the basis for the division into the following user groups:

The user groups provide the basis for the following personas.

Table 2: Eurostats distribution of individuals using the Internet for reading and downloading online newspapers and news magazines.

Description	%
<i>student</i>	35
<i>employees, self-employed, family workers</i>	31
unemployed	22
retired or other inactive	13
Description	%
<i>high formal education</i>	50
medium formal education	33
no or low formal education	17
Description	%
<i>male</i>	55
female	45
Description	%
<i>16-24 of age</i>	34
<i>25-54 of age</i>	34
<i>55-64 of age</i>	21
<i>65-74 of age</i>	11

A.1.1 Thomas: student medium formal education male of age 21

Thomas is 21 and a student at the Technical University of Denmark to be a bachelor of engineering in software. He is very interested in soccer and is therefore always updated on sports news. He reads about it online, newspapers and talks about it with friends. With big events he even likes to post it on Facebook. As a soon-to-be software engineer he has a natural thirst for news about technology, and he mainly reads these at home at the dormitory. wired.com, newz.dk, engadget.com, facebook.com computer, Samsung Galaxy Tab

A.1.2 Laura: employed high formal education female of age 39

Laura is 39 and is employed as a key account manager. She likes to be updated on strategies and economical status of rivalling companies. She is also very interested in politics and likes to discuss this subject with her friends. She reads economical news and likes to be updated on the run. b.dk, borsen.dk, twitter.com iPhone, iPad

A.1.3 Marie: unemployed no or low formal education female of age 61

Marie is 61 and a currently unemployed housekeeper. She spends her day looking for a job and taking care of her pet cat until her husband comes home. She mostly looks for the gossip sections or news about crime or big disasters. She also spends some time reading through the travelling guides as she dreams of going away with her husband. ekstrabladet.dk, bt.dk, nyhederne.tv2.dk computer, Lenovo IdeaPad A1

A.1.4 Carl: retired or other inactive high formal education male of age 69

Carl is a retired professor in psychology. He likes to discuss human behaviour and relation with his acquaintances and is very interested in cultural events. Therefore he often seeks the cultural sections and discussion fora to see what is going on. politiken.dk, aok.dk, dr.dk computer, iPad

A.2 SCENARIOS

A.2.1 Thomas

Thomas comes home after a day at the study, picks up his tablet computer and opens Editor from the desktop. Editor opens and shows him the front page where all the headlines stories are displayed. The main story is about a new version of the Android OS that has been released today and presses it to read more. The story opens in a full window display with quality images to match the articles. He reads the first section and feels satisfied with the amount of information, but wants to share the information on Facebook, so he clicks share button and writes a comment and posts it on his Facebook wall. He closes the article and returns to the front page. He sees a top story below the main story about Mr. Mærsk Mc-Kinney Møller who has died. It is not a story that falls into his key interests, but as the news is big he is satisfied that he got informed about it. Thomas feels like reading more about technology so he opens the menu and chooses the "Tech" section he has installed in the application. The section opens with a head line and a page number to let him know where in his paper he has navigated to and finds an article about a new multicore CPU technology. He has never been interested in CPU technology before, but finds this technology interesting after reading about it, so he opens the application settings and types in keywords about the technology under his "Tech" section to keep him updated about it. He also adjusts the ratio between general and personal news, to be less personal as he feels like he needs to broaden his horizon a bit with respect to news. He closes the settings menu

and Editor immediately starts updating the articles. Some new articles about CPU technology has been included amongst the articles in the “Tech” section after paging through the section and reading some of the most interesting articles he closes the application.

It could be nice if the key words of a story could be or is already highlighted, so he can click it and add it to his positive or negative list.

A.2.2 *Laura*

Laura is on the train on her way to a business meeting this morning and pulls out her tablet and sees she has one notification from Editor. She opens Editor to get updated on todays news. The front page is displayed and there are headlines from different top articles and a notification is shown in the corner. She presses the notification and the pages turns to show her the article, which opens in full screen. After reading it she wants to see todays headlines, so she presses the back button to return to the paper and presses the return to front page button and the paper turns pages to reach the front page. She scans the page to see if there is any big news about her rivalling companies. There is no breaking news, so she just turns the page to browse the content of todays paper. As she browses the “Politics” section of her paper she finds an article about the Prime Minister introducing a new bill about a toll ring around the capitol city. She chooses the article and it is shown in full screen. As she reaches the bottom of the article she sees the comments about it where her friends and most others are against it. She decides to join the discussion and posts a comment on the article wall. She also sees one of her friends has not commented on the article wall and decides to share the article with her as she thinks she would agree with her opinion. She presses the share button and chooses the Editor logo. A list of her friends is shown, some of them who has already read the article is greyed out, but the one she was looking for is not. So she chooses her and a notification is sent to her.

A.2.3 *Marie*

It is morning and Marie wants to check the news with her coffee in the couch, so she opens Editor from her tablet to get updated. The front page is displayed with a collection of stories as highlights of the content of the paper. It mainly contains stories about celebrities and a big disaster that has happened in Japan, but there is also a story about a big political change, that she does not find interesting. So she goes to the settings menu and types in "politics" to add to her negative list. She also adjusts the personal/general news ratio to contain only personal news as she wants only news that is directed to her. She returns to the front page which is now free of political stories. Her newspaper contains many images and videos as she has set her graphical/textual content ratio more towards graphical content.

A.2.4 *Carl*

Sunday morning Carl wakes up, puts over the kettle to make a cup of coffee. While he waits for the water to boil he picks up his iPad and opens Editor to check the news. The front page opens with headlines from the different sections. There is a review article about a new show in the theatre. Carl presses the article and the system turns pages to the "Cultural" section of his newspaper and opens the article in full screen. Because the show gets good ratings he decides to order some tickets to him and his wife, which he does using the devices browser. After this he reopens Editor which opens in a display of the same article, as he left it. Carl pours his coffee and turn to the "Funnies" section. The section displays snippets of comics strips and some humorous and opinionated articles.

A.3 BUSINESS CASE

A.3.1 *Need*

User value: personal quality and up-to-date stories enriched with quality images. This means that content providers should be chosen/verified. Same navigation as actual newspapers, but faster and with endless more content. Instantly up-to-date. Adaptive layout. Adjustable user profile.

A.3.2 *Approach*

personalised content + composition.

Constraint Programming: fast computation - good for optimal solutions, describes the generic solution instead of how to solve or find it, very easy to tailor the problem definition of the solution and adjust it and even let users make the adjustments - transparency.

Content providers can get to know their readers preferences better and improve the provided content.

A.3.3 *Benefit Per Cost*

Revenue flow: Content providers are paid. Income from advertisers (scattered [[Ovesson and Wikström, 2005](#), p. 6-7]) and users. Income from selling user behaviour patterns and targeted commercials.

A.3.4 *Competition*

FlipBoard, Wired magazine, Zite and app with actual editors affiliated.

A.4 REQUIREMENTS

The above scenarios, user needs and business case led to the following requirements.

A.4.1 *Non-functional Requirements*

- “the clear overview of content, including a beginning and an end, the ease of use, typography and design” [Ihlström *et al.*, 2004, p. 7]
- both general and personal news (collaborate filtering solves that some news are not received, but are universally interesting [Díaz and Gervs, 2005])
- familiarity in design from printed paper [Ihlström *et al.*, 2004, p. 7]
- Design and layout from printed newspaper [Åkesson *et al.*, 2005]
- both images and videos - test
- a good ratio of graphical and textual - test
- front page should give a good overview of the content - test
- “news valuation, e.g. positioning of lead story” [Ihlström *et al.*, 2004, p. 7]
- mobility [Ihlström *et al.*, 2004, p. 7]
- continuous updates [Ihlström *et al.*, 2004, p. 7]
- “easy and intuitive navigation” [Ihlström *et al.*, 2004, p. 7]
- add video and sound [Ihlström *et al.*, 2004, p. 7]
- incorporate social community and social networks

A.4.2 Calculations on number of columns

iPad screen size:

$197 \times 148mm$ or $1024 \times 768px$

International Herald Tribune (the global edition of the new york times): $\frac{398mm}{6} = 66.33333333mm$

Børsen (uses both 5 and 6 columns): $\frac{285mm}{5} = 57mm$ and $\frac{285mm}{6} = 47.5mm$

Information (5 columns 4 on the back): $\frac{285mm}{5} = 57mm$ (back $\frac{285mm}{4} = 71.25mm$)

Guardian (5 columns): $\frac{314mm}{5} = 62.8mm$

Politiken (6 columns): $\frac{392mm}{6} = 65.33mm$

Berlingske (4 columns): $\frac{285mm}{4} = 71.25mm$

Average on the most regular columns:

$\frac{66.3+57+57+62.8+65.3+71.3}{6} = 63.28333333mm$, i.e. 3.11 columns in landscape and 2.34 in portrait.

(Average on every column width:

$\frac{66.3+57+47.5+57+71.25+62.8+65.3+71.3}{8} = 62.30625mm$, i.e. 3.16 columns in landscape and 2.38 in portrait.)

1200px screen:

$\frac{1200 \cdot \frac{197}{1024}}{63.28} = 3.65$ columns, where $197/1024$ is px to mm ratio and 63.28 is the thinnest column width

column size in px:

$$63.28 \cdot \frac{1024}{197} = 329px$$

$$62.31 \cdot \frac{1024}{197} = 324px$$

$$\text{average} = 326px$$

A.4.3 Functional Requirements

- 2 columns in portrait and 3 in landscape
- “open, turn pages, chose article, read and return” [Ihlström *et al.*, 2004, p. 6]
- section headlines [Ovesson and Wikström, 2005, p. 6-7]
- article headlines
- article summaries / extracts [Díaz and Gervs, 2005]
- menu w. section headlines [Ovesson and Wikström, 2005, p. 8]
- page numbers [Ovesson and Wikström, 2005, p. 6-7]
- press “like” or key word based user profile (mark self or highlighted? right click to add): positive + negative list (keywords+categories [Abuzir and Vandamme, 2002], [Díaz and Gervs, 2005] and [de Buenaga Rodríguez *et al.*, 2004])
- full screen display of article
- organise into personalised sections
- opens in front page view (summery of newspaper 8 articles) [Ovesson and Wikström, 2005, p. 8]
- adjust variables
- share directly (grey out the ones who have read it)
- comment
- see friends comments
- “The presentation schema – headline, abstract, and text, together with a relevance value with respect to the user profile – rates the highest in terms of user satisfaction, and yet it is not the most frequent.” [Díaz *et al.*, 2001]

- ability to search [Ihlström *et al.*, 2004, p. 7]
- Landscape + portrait [Ovesson and Wikström, 2005, p. 6-7]
- touch screen interaction [Ovesson and Wikström, 2005, p. 6-7]
- Functionality from online newspaper [Åkesson *et al.*, 2005]
- Name of columnist [de Buenaga Rodríguez *et al.*, 2004, p. 4]
- Transparency of implicit relevance feedback (see/modify current weights of categories) [de Buenaga Rodríguez *et al.*, 2004, p. 7]
- dynamic short-term + static long-term user profile [Abuzir and Vandamme, 2002], [Díaz and Gervs, 2005] and [de Buenaga Rodríguez *et al.*, 2004]
- relevance feedback [Abuzir and Vandamme, 2002], [Díaz and Gervs, 2005] and [de Buenaga Rodríguez *et al.*, 2004]

A.5 TEST RESULTS

This section sums up the test results in an unordered list.

- Touch friendly interaction
- Tools for changing the layout, like changing the font size and colour scheme
- The front page should give an overview
- View whole menu all the time
- Give suggestions to similar articles, i.e. more on this story, subject and topic

- List overview of headlines in top of sections
- Search within relevant articles. Search bar should be visible at all times.
- Indication of similarity on articles
- Archive possibility
- User feedback notated with “relevant” and “irrelevant”
- White space besides articles is not a problem
- General layout corrections
- Better visual division between articles
- More and larger images
- No need for general news, personalised news is enough
- Choose categories with topics to get started
- Ability to choose period show articles from
- Ability to choose when the newspaper should be generated, e.g. on Fridays to be read in the weekend
- Dividing columns into screens
- Social community implementation
- Visualisation of user behaviour
- Possibility to use it for research
- Get articles from magazine and newspaper subscriptions, e.g. by adding them to a specific section
- It has solved the problems that [http://nyhederne.tv2.
dk/](http://nyhederne.tv2.dk/)¹ has, i.e. confusing layout, no overview and hard to navigate

¹ The website of a Danish news channel.

BIBLIOGRAPHY

- [Abidi and Chong, 2004] Syed Sibte Raza Abidi and Yong Han Chong. Constraint satisfaction methods for information personalization. Technical report, Dalhousie University, Canada and Universiti Sains Malaysia, Malaysia, 2004.
- [Abuzir and Vandamme, 2002] Yousef Abuzir and Fernand Vandamme. E-newspaper classification and distribution based on user profiles and thesaurus. Technical report, University of Ghent, 2002.
- [Apt, 2006] Krzysztof R. Apt. *Principles of Constraint Programming*. Cambridge University Press, 2006.
- [Billsus and Pazzani, 2000] Daniel Billsus and Michael J. Pazzani. User modeling for adaptive news access. pages 147–180, 2000.
- [Bird *et al.*, 2009] Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, Inc., 2009.
- [Bouras and Tsogkas, 2010] Christos Bouras and Vassilis Tsogkas. W-kmeans: Clustering news articles using wordnet. pages 379–388. Springer Berlin / Heidelberg, 2010.
- [Bush, 1945] Vannevar Bush. As we may think. *The Atlantic Monthly*, 1945.
- [Centeno *et al.*, 1999] Vicente Luque Centeno, M^a Carmen Fernández Panadero, Carlos Delgado Kloos, Andrés Marín López, and Carlos García Rubio. Personalizing your electronic newspaper. Technical report, Universidad Carlos III de Madrid, 1999.
- [Claypool *et al.*, 1999] Mark Claypool, Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Nletes, and Matthew Sartin. Combining content-based and collaborate filters in an online

newspaper. Technical report, Worcester Polytechnic Institute, Massachusetts, 1999.

[de Buenaga Rodríguez *et al.*, 2004] Manuel de Buenaga Rodríguez, Manuel J. Maña López, Alberto Díaz Esteban, and Pablo Gervás Gómez-Navarro. A user model based on content analysis for the intelligent personalization of a news service. Technical report, Halmstad University, 2004.

[Díaz *et al.*, 2001] Alberto Diaz, Pablo Gervas, Antonio Garcia, and Inmaculada Chacon. Sections, categories and keywords as interest specification tools for personalised news services. *Online Information Review, OIR*, 2001.

[Díaz and Gervs, 2005] Alberto Díaz and Pablo Gervs. Personalisation in news delivery systems: Item summarization and multi-tier item selection using relevance feedback. Technical report, Universidad Complutense Madrid, 2005.

[Eirinaki and Vazirgiannis, 2003] Magdalini Eirinaki and Michalis Vazirgiannis. Web mining for web personalization. Technical report, Athens University of Economics and Business, 2003.

[Esteban *et al.*, 2000] Alberto Esteban, Pablo Gómez-Navarro, and Antonio Jiménez. Evaluating a user-model based personalisation architecture for digital news services. In *Research and Advanced Technology for Digital Libraries*, pages 259–268. 2000.

[Esteban *et al.*, 2001] Alberto Díaz Esteban, Manuel J. Maña López, Manuel de Buenaga Rodríguez, Jose María Gómez Hidalgo, and Pablo Gervás Gómez-Navarro. Using linear classifiers in the integration of user modeling and text content analysis in the personalization of a web-based spanish news service. Technical report, Universidad Europea de Madrid, Universidad de Vigo and Universidad Complutense de Madrid, Spain, 2001.

[Genette, 1997] G. Genette. *Paratexts: Thresholds of interpretation*. Number 20. Cambridge Univ Pr, 1997.

- [Haskins, 1965] Jack B. Haskins. The editorial mix: One solution to a magazine editor's dilemma. Technical report, Journalism & Mass Communication Quarterly, 1965.
- [Holmqvist *et al.*, 2003] K. Holmqvist, J. Holsanova, M. Barthelson, and D. Lundqvist. Reading or scanning? a study of newspaper and net paper reading. *Mind*, 2(3):4, 2003.
- [Holsanova *et al.*, 2006] J. Holsanova, H. Rahm, and K. Holmqvist. Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements. *Visual communication*, pages 65–93, 2006.
- [Ihlström *et al.*, 2004] Carina Ihlström, Maria Åkesson, and Stig Nordqvist. From print to web to e-paper - the challenge of designing the e-newspaper. Technical report, Halmstad University, 2004.
- [Kress and others, 1999] G. Kress et al. *Reading images*. Routledge, 1999.
- [Liu *et al.*, 2001] Fei Liu, Yupin Luo, and Masataka Yoshi-kawaf Dongcheng Hu. A new component based algorithm for newspaper layout analysis. Technical report, Tsinghua University, Beijing, 2001.
- [Lloyd, 1994] JW Lloyd. Practical advantages of declarative programming. Technical report, University of Bristol, 1994.
- [Mobasher, 2007] Bamshad Mobasher. Data mining for web personalization. In *The Adaptive Web*, Lecture Notes in Computer Science, pages 90–135. Springer Berlin / Heidelberg, 2007.
- [Nielsen, 2012] Jakob Nielsen. How many test users in a usability study? <http://www.useit.com/alertbox/number-of-test-users.html>, 2012.
- [Nilsson, 1999] Jørgen Fischer Nilsson. A conceptual space logic. Technical report, Department of Information Technology Technical University of Denmark, 1999.
- [Ovesson and Wikström, 2005] Fredrik Ovesson and Kristin Wikström. From visions to specification - using user de-

signed mock-ups for envisioning user requirements for the future e-newspaper. Technical report, Halmstad University, 2005.

[Perkowitz and Etzioni, 2000] Mike Perkowitz and Oren Etzioni. Adaptive web sites. *Commun. ACM*, pages 152–158, 2000.

[Russell and Norvig, 2003] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd edition, 2003.

[Tidwell, 2010] J. Tidwell. *Designing Interfaces*. O'Reilly Media, Inc., second edition, 2010.

[Vossen, 2005] M.P.H. Vossen. Local search for automatic playlist generation. Master's thesis, Technische Universiteit Eindhoven, 2005.

[Åkesson *et al.*, 2005] Maria Åkesson, Carina Ihlström, and Jesper Svensson. How would you like your e-newspaper? – converging the best from two worlds. Technical report, Halmstad University, 2005.