2nd International Workshop on Web Document Analysis (WDA), Edinburgh (Scotland), 2003

Dynamic Generation of Multi-Modal Crosswords in Web Documents

Stefan Jaeger, Masaki Nakagawa
Tokyo University of Agri. & Tech.
Department of Computer Science
2-24-16 Naka-cho, Koganei-shi
Tokyo 184-8588, Japan
stefan@hands.ei.tuat.ac.jp

Hermann Hild, Roald Wolff, Guido Wojke
SMI Cognitive Software GmbH
Haid-und-Neu-Str. 7
76131 Karlsruhe, Germany
hild@smi-software.de

Abstract

Solving crosswords in newspapers and journals has become a popular pastime for myriads of readers. Accordingly, many publishers offer crosswords to their readers on a regular basis. With the print media going more and more on-line in the Internet, some publishers have started to transfer the classic paper-based crossword to their online edition in order to attract more people to their web pages. This paper describes the problems of this transfer and our solution exploiting the possibilities offered by electronic web documents, including handwriting recognition and customization.

1. Introduction

Web documents offer many advantages over standard paper documents. Two of the most important advantages are perhaps multi-modal interaction and customization: Multimodal interaction allows users to access and enter more information in a much faster way, and with less cognitive load. Customizable documents can change their structure dynamically and are thus able to present information in a way that fits best the needs and requirements of individual users.

Nevertheless, one should keep in mind the striking advantages of standard paper, especially in combination with a pen, such as ease of use, low weight, low price, resolution, etc. In fact, care must be taken when transforming classic paper documents into web documents as the lack of these features may easily negate the advantages of web documents. In this respect, web documents need to be carefully designed in order to become widely accepted alternatives to paper documents or replacements of paper-based work flows.

Crosswords are a good example here. To realize the full potential of web-based crosswords and to make them real alternatives to paper-based ones, support of pen input is, of course, very important. This paper describes our implementation of classic crosswords as parts of web documents. We provide writers with the same look and feel they normally enjoy with paper plus additional features such as handwriting recognition and learning capability.

The structure of our paper is as follows: Section 2 discusses typical features of crosswords with regard to their integration into web documents. Section 3 describes the interactive features we augmented crosswords with to realize their full potential in web documents. The processing of pen-input, which is of course the most natural way of entering data into a crossword, is described in Section 4. Dynamic content generation and customization of crosswords is presented in Section 5. Section 6 discusses crosswords in the more general context of web documents and client-server applications. Finally, a summary concludes this paper.

The work described in this paper is protected under German Patent 20014722 (see Reference [1] for more details).

2. Crosswords

Crosswords are very popular games that can be found in many newspapers nowadays. One reason for their popularity is surely the small number of requisites needed to play a game: Crosswords merely require a pen.

There are different types of crosswords, each putting emphasis on a slightly different aspect of the game. While the popularity of each type varies among countries and users, there is basically one key principle they all have in common: Crosswords provide clues to unknown words that the player has to guess right. In doing so the player may use letters of words he or she already guessed right, which serve as additional clues to still unknown words. We have focused our work on a specific, widespread crossword type. This type is characterized by a rectangular character grid with

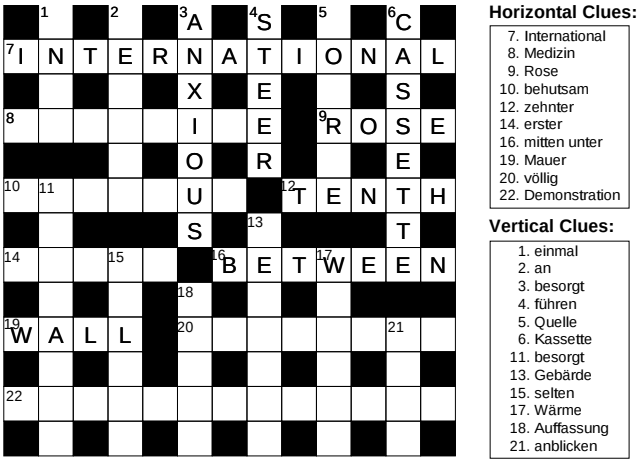

Figure 1. Crossword example.

clues printed either directly into the grid or on its side (see Figure 1). Nevertheless, virtually all the issues presented and discussed in this paper should be easily transferable to other crossword types.

Pen-based applications, such as crosswords, are inherently two-dimensional, accepting graphical input usually written onto an even plane. Among them, however, there are many that in some sense are "more two-dimensional" than others. These applications are characterized by extensive pen-up and pen-down movements in both the vertical and horizontal direction, exceeding the number normally observed in general handwriting. Typical examples of those applications are, for instance, generation of mathematical formulas and form filling. Mathematical formulas are composed of various syntactical elements and structures impeding a continuous writing from left to right, such as exponents, indexes, fractions, and many more. The same can be observed for forms: Writers very often do not fill out forms sequentially but instead enter data arbitrarily into form fields, especially when they are uncertain about the meaning of a form field or just do not know the information required. Naturally, web documents containing those "highly two-dimensional" applications are best processed on pen devices, such as Tablet PCs or other graphical devices supporting pen-input. A standard mouse and keyboard interface to these applications provides only an unsatisfactory and thus sub-optimal way of interacting with web documents here (see current formula editors). In particular, the continuous switch from cursor positioning to text entry and back puts additional strain on the writer. Crosswords are no exception in this respect. In fact, leaping to entry fields all across the crossword is part of the game.

Several publishers have tried to enhance their web pages by adding on-line crosswords, with modest success so far. The interfaces they have come up with do not take into account the specific requirements of crosswords. Their solutions adhere to the standard mouse-keyboard interface and do not consider the fact that pen-input is an integral part of crosswords. For instance, most interfaces require the user to press the space bar in order to switch from vertical writing mode to horizontal writing mode and vice versa. This enables the cursor to automatically jump to the next input field once a character has been entered via the keyboard. Nevertheless, entering words usually requires a repositioning of the cursor by moving and clicking with the mouse. Needless to say that those unnatural user interfaces have prevented on-line crosswords from enjoying the same great popularity as their paper-based cousins.

Fortunately, pen-input is nicely supported by many of today's hardware devices, such as Tablet PCs. For these devices, we designed an interface that allows users to solve web-based crosswords in the same way they would solve them on paper, avoiding all the problems mentioned above. The following sections describe important aspects of our system, including additional interactive features beyond those offered by classic paper-based crosswords.

3. Interactivity

Crosswords are inextricably linked to handwritten characters. Pen and paper provide an incredible spartan but functional interface that allows writers to take so much pleasure in simple crosswords. Therefore computer aided crosswords should support pen input of handwritten characters to not spoil the joy users have in this game. While support of mere pen-input is already sufficient for a decent computer aided crossword system, this is only the halfway point on our way to an even more powerful interface: Pen input suggests handwriting recognition, which can enhance the crossword interface considerably and enable interactions not possible with paper [3]. Handwriting recognition, in combination with gesture recognition, allows more sophisticated customized user interaction, such as learning capabilities and context-dependent clues. For instance, we implemented a gesture recognizer that recognizes question marks written into a character box of a crossword. By writing a question mark into a character box, the user can ask for a hint at or solution to either a single character or a whole word, which are then displayed on the screen.

Pen input is, of course, the most natural form of entering data into an (electronic) crossword. Nevertheless, we also support to some extent speech input, though we do not expect speech to become the main input medium. We trained a spelling recognizer based on a multi-state time-delay neural network and integrated it in our electronic crossword [2].

This enables the user to utter individual letters instead of writing them. In our current implementation, spelled characters appear in a highlighted cursor field, which automatically proceeds to the next empty field upon each entry.

Another issue that needs to be addressed when implementing interactive crosswords is the cursor problem. A plain mouse-and-keyboard interface depends on a cursor that shows the user where his input will appear in the crossword. The user can then reposition the cursor to a new input field by moving the mouse. A pen interface does not need such a cursor per se since the position of the pen tip implicitly indicates the input position in the crossword. It can nevertheless be very advantageous sometimes to have a cursor implemented also in pen interfaces. For instance, a cursor allows handwritten input anywhere on the screen, not only above the intended input field. Moreover, a cursor enables cursively written words as input into vertical columns. A speech interface also clearly benefits from a cursor-based input since addressing input fields by speech is highly impracticable. Despite these advantages, we have tried to avoid a cursor wherever possible. The reason is that cursors have no counterpart on paper. They would thus impose a somewhat unfamiliar and unnatural input technique on the user. The abandonment of the cursor concept requires us to automatically assign input strokes to their correct character fields though. In our current implementation, we still offer the cursor concept as an option for the user, mainly to support position-independent entry of characters.

4. Handwriting Recognition

Handwriting recognition in general is still subject of intensive study, and the reliable recognition of freely and cursively handwritten text is still an open problem. Fortunately, crosswords are not very demanding in terms of handwriting recognition. The standard crossword expects discrete, pre-segmented letters and provides a well-defined box for each of them. The current state-of-the-art in handwriting recognition can manage this simplified recognition problem and provide satisfying recognition rates, also for writer-independent recognition [3].

The recognition problem becomes more difficult when we allow not only single characters but also cursively written input. However, this is a rather untypical input that usually does not occur in paper crosswords. We also support this type of handwriting but, like speech, do not expect it to become a significant alternative to single character input.

Our symbolic gesture recognizer is a refined dynamic time warper that allows users to define their own symbols. However, there is a small set of predefined symbols (e.g. a question mark) allowing users to ask for clues to unknown letters or words.

As was said above, we try to avoid a somewhat unnatural cursor concept and want to adhere strictly to the natural pen-and-paper interface. This leaves us with the problem of assigning each stroke of a handwritten on-line trajectory to the correct character field. We formulate this problem as an optimization problem and solve it with standard methods. Our objective function is a combination of classifier likelihood and geometrical attributes. For each combination of neighboring strokes the handwriting recognizer computes a list of recognition results together with their likelihood values. These values are combined with a geometrical attribute, which is computed for each character field overlapped by the stroke combination. This attribute is a spatial measure describing how the stroke combination fits a specific character field. The search for both the optimal stroke partition and the best assignment of partitioned strokes to character fields is then organized as a search for the optimal path in a directed acyclic graph. Thanks to dynamic programming, this becomes a linear search.

5. Dynamic Content Generation

The capability to generate dynamic content is one of the big advantages electronic crosswords have over crosswords in the print media. An electronic crossword system can, for instance, offer crosswords composed of vocabularies specified either by the crossword maker or writer himself. For this purpose we developed an automatic crossword generator enabling a user to compile his or her own crosswords.

Dynamic crossword generation is a typical search problem in Artificial Intelligence. The objective is to accommodate as many words as possible in a crossword grid of given size. Additional constrains are often used to help confine the search space, e.g. maximum and minimum word lengths etc. We must admit, however, that the quality of completely automatically generated crosswords does not seem to match the one of classic paper crosswords. The quality of a crossword, in this context, is defined by its number of remaining black squares; i.e. the number of squares that do not absorb characters and thus cannot be filled out by the writer. The less black squares a crossword contains, the more it will be appreciated by the user. To the best of our knowledge, most professional crossword makers use a semi-automatic approach to crossword generation. They first generate a computer-aided, raw skeleton which is then further elaborated by hand. Another problem in automatic crossword generation are clues, which are hints at the correct solutions given to the user. Clues are usually printed either directly into the crossword or beside the puzzle, where they are divided into clues to vertical words and clues to horizontal words. Clues are utmost important since they represent the character of a crossword. They are therefore directly related to user satisfaction. Unfortunately, clues require creativity

and ingenuity, both features that modern Artificial Intelligence still has difficulty coping with. Clues are thus very difficult to automate. This is the reason why we confined ourselves to a simple database of pre-compiled clue alternatives that are chosen according to requirements.

Dynamic content generation is tightly connected with the crossword's learning capabilities: Electronic crosswords are able to evaluate the performance of a user on a given vocabulary and adapt their behavior and content accordingly, which is a feature not realizable with conventional paper crosswords. This enables electronic crosswords to serve as valuable learning tools teaching, for instance, foreign vocabularies. Learning vocabularies by solving crosswords is an enjoyable and thus more effective alternative to what would otherwise be a tiresome and monotonous learning process. A fact that should make crosswords even more popular than they already are.

There are several ways of evaluating the performance a user achieves on a generated crossword. We use an evaluation scheme that is based on how certain or confident a user is of his answers. For each answer to a word clue, we evaluate the user's confidence in terms of four parameters:

1. the number of previous, incorrect guesses
2. the number of letters known due to other word guesses
3. the help provided by the system (disclosed letters, ...)
4. the number of words already guessed

According to these parameters, we identify the words the user has still problems with and compile a new crossword biased in favor of these words. This technique guarantees a continuous stream of crosswords that are both new and challenging for the user. By steadily adding unknown words to the vocabulary, we create a progressive learning process that is well adapted to the user's learning curve. In particular, our learning component consist of a static and a dynamic part. The static part is basically a database containing pairs of clues and their corresponding answers. The dynamic part contains writer-specific information and is therefore different for each writer. Typical writer-specific informations are, for instance: performance evaluations, game scores, aborted games, and suchlike.

6. Crosswords in Web Documents

Accessing crosswords via the Internet poses no major technical problem. Crosswords can be easily integrated into existing web pages. There are actually many websites in the Internet that contain crosswords integrated as Java Applets. However, none of them has supported handwritten input or dynamically generated content so far.

Crossword software can be implemented either as a self-contained application running on the user's computer or as a client-server architecture. The latter option is more flexible in the sense that the server can supply the client with a continuous stream of new crossword content. A client-server architecture is also better suited to customer care: A weekly downloadable crossword contest, for instance, could be a typical customer service offered by a content provider. In terms of handwriting recognition, a client-server architecture can release the client from any handwriting recognition tasks and put the responsibility for handwriting recognition solely on the server, enabling handwriting recognition also on the smallest, portable pen-computer. This is an important point, especially when considering the demanding hardware and software requirements of today's handwriting recognizers, though the requirements for single character recognition in crosswords can be considered moderate in this respect.

We have developed a client-server system that transmits handwriting from the client to the server, recognizes it on the server, and sends the recognition result back to the client. However, we have not yet coupled it with our crossword software.

7. Summary

In this paper we presented a rough description of our electronic crossword system. This system accepts handwriting input as well as speech input. It is capable of generating dynamic crossword content for predefined vocabularies. Which, in combination with dynamic performance tracking, provides a nice platform for learning vocabularies of foreign languages.

We hope that our system will not only become an adequate electronic alternative to paper crosswords, but that its additional features will also attract new devotees of crosswords. Our system can run both as a stand-alone application or as an extra service enhancing normal websites. The latter may result in higher hit rates and more user satisfaction.

References

[1] H. Hild and S. Jaeger. Verfahren und Vorrichtung fuer ein elektronisches Kreuzwortraetsel. German Patent 20014722, 2002 (in German).

[2] H. Hild and A. Waibel. Speaker-Independent Connected Letter Recognition with a Multi-State Time Delay Neural Network. In *3rd European Conference on Speech, Communication and Technology (EUROSPEECH)*, volume 2, pages 1481–1484, Berlin, 1993.

[3] S. Jaeger, S. Manke, J. Reichert, and A. Waibel. Online Handwriting Recognition: The Npen++ Recognizer. *International Journal on Document Analysis and Recognition*, 3(3):169–180, 2001.