KANJIRU: A SOFTWARE INTERFACE FOR VISUAL KANJI SEARCH
Chris Winstead and Atsuko Neely, Utah State University

Abstract: A new software interface, called "Kanjiru," provides a user-adaptive visual interface for searching Kanji characters. Users navigate through a hierarchy of basic stroke patterns to locate complex characters, words and their associated definitions. The interface uses a statistical algorithm to estimate the user's most probable target character. The statistical algorithm makes Kanjiru an intuitive and mistake-tolerant interface, enabling users to identify Japanese characters and words with only a minimal understanding of Kanji.

Keyword: Kanji computer dictionary, adaptive user interface.

INTRODUCTION

Kanjiru is a game-like computer interface for searching Kanji characters based on visual appearance. With Kanjiru, a user can quickly find a character by "gliding" across basic stroke patterns. Unlike conventional dictionaries, Kanjiru can identify characters based on any visual feature that the user recognizes. Users do not need to understand stroke order or the radical system to successfully find characters and words using Kanjiru.

The Kanjiru interface, shown in Figure 1, uses a "gliding" interface similar to David MacKay's accessibility interface called "Dasher" (Ward, 2002). When a search begins, the user is presented with a collection of colored boxes that contain basic patterns. The user can "zoom in" on a box by moving the mouse toward it. The user then appears to "fly toward" the box and the pattern it contains. As the user zooms closer, a new layer of patterns appears within the box. Each new pattern in the new layer is more complex than those in the previous layer. Descriptions therefore evolve, beginning with simple morphology and advancing into exact patterns.

The user glides through the Kanjiru interface to locate patterns associated with a particular target character. The ultimate goal is to locate the target character in a detailed Kanji dictionary. As the user glides across layers of boxes, Kanjiru continually updates a match list and displays it to the user. Once the user's target character appears in the match list, the search is finished.

Even individuals with no prior knowledge of Kanji are able to quickly learn and use the Kanjiru interface. After a few hours of experience with Kanjiru, users can locate characters typically in well under a minute. Once the desired character is found, the user can click to access a dictionary to view words, pronunciations and definitions associated with the character. Currently, the interface links to pages from Jim Breen's online Japanese dictionary server (Breen, 2007).

Current Kanji dictionaries require considerable user knowledge to be used effectively. Even intermediate students of Japanese easily make mistakes, forcing them to repeat their searches. Kanjiru is a *mistake-tolerant* interface. There is no single correct classification for a character. Instead, the interface uses a statistical algorithm to guess which characters the user is *most probably* searching for, based on the user's evolving description. This list is continually updated as the user searches, creating a
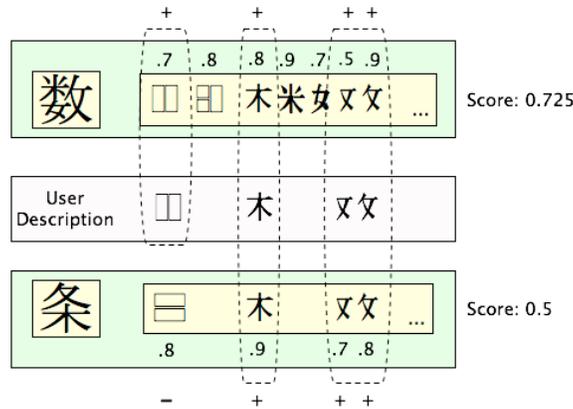
Main Window   Patterns   Match List   Dictionary View

Figure 1: The Kanjiru interface.

short list of characters related by visual similarity. The statistical matching algorithm frequently identifies the desired character based on only one or two pieces of information.

HOW KANJIRU WORKS

Kanjiru is an adaptive interface that is trained to recognize characters based on accumulated user perceptions. In training mode, users are asked to describe randomly selected characters. A description is created when the user glides through Kanjiru and selects patterns that appear to belong to the character. After several descriptions are collected for a character, a weighted *average description* is created. If there are N total descriptions for a character, of which M descriptions include pattern x, then the average description contains x with a weight $w(x)=M/N$.

During normal use, as the user glides across patterns in Kanjiru, a *user description*, D, is created. The description contains a collection of L patterns, x_j, where $1 \leq j \leq L$. Each *character* C_i in the database has an *average description*, A_i, which represents user information collected through training. A *score* S_i is assigned to each C_i as follows. Let $w_i(x)$ be the weight assigned to pattern x in description A_i. If a pattern x_j in D is also in C_i, then we add $w_i(x_j)/L$ to S_i. If a pattern x_j in D is not in A_i, then we subtract a *penalty P* from S_i. The highest-scoring characters are presented to the user in the match list. This procedure is illustrated for two characters in Figure 2.

EVALUATION AND USER EXPERIENCES

Kanjiru is intended to provide an easily learned interface that allows even total novices to begin studying Kanji. A total novice is a user who has no prior knowledge of Kanji or the Japanese language. To evaluate Kanjiru's effectiveness, two experiments

Figure 2: How Kanjiru calculates character scores from user descriptions. In this example, the penalty is *P* = 0.1.

were performed. In the first, a group of total novices was asked to search for randomly selected characters using Kanjiru. The goal of this experiment was to measure how quickly novices could learn and use the Kanjiru interface. In the second experiment, students with prior Kanji training were asked to find the same randomly chosen characters in Kanjiru and using a standard Kanji dictionary.

The first experiment focuses on measuring *learning rate* and *failure rate.* The learning rate of Kanjiru users is estimated by measuring the average search time vs the number of search trials the user has conducted using the interface. If a character is not found within ten minutes, users are instructed to record the search as a failure and move on to a new character. The average search time and failure rate are expected to decrease as users gain experience using Kanjiru.

Results of the first experiment are shown in Figure 3. When users first encounter the interface, the average search time and failure rate are roughly 230 seconds and 53%, respectively. After 25 trials, these figures are reduced to 50 seconds and 12%, respectively. The novice users were able to find arbitrary characters in less than one minute, with a high probability of success, after one to two hours of using Kanjiru.

The second experiment focuses on comparing Kanjiru to traditional dictionaries. In this experiment, the participants have roughly one semester of experience studying Kanji. Each student is asked to locate a randomly chosen character using Kanjiru, and then to find the same character in the traditional dictionary. Both searches are timed.

Figure 4 shows the results from two students who participated in the second experiment. One student understood Kanjiru immediately, and was able to find characters 37.7% faster using Kanjiru. The other student needed some time to learn the Kanjiru interface. Initially, this student needed an average of 107 seconds longer to find characters using Kanjiru. By the end of the session, the student found characters an average of 31.8 seconds faster using Kanjiru, 47.8% faster than the student's average dictionary search time.

At present, only a few students have participated in the second experiment, and data collection is ongoing. One student offered a simple explanation for the lack of volunteers: "the dictionary is boring." This remark highlights an important attribute of the Kanjiru interface: it keeps the students' attention. Volunteers are happy to spend hours gliding through Kanjiru's game-like environment, but are generally unenthusiastic about

dictionary work, and only participate when course credit is offered. This provides qualitative evidence that Kanjiru may be an effective way to maintain students' motivation by making Kanji a less tedious subject.
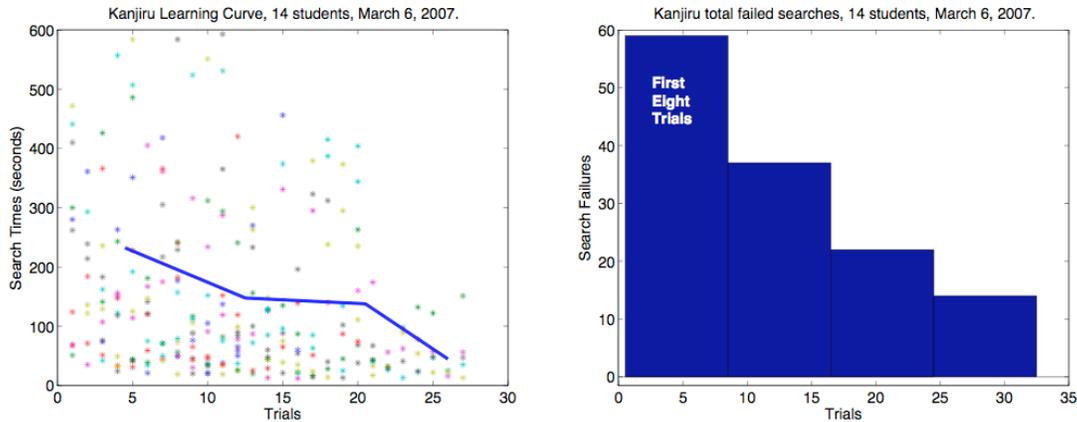

Figure 3: Learning curve and failure curve for a group of total novices. Students had no prior exposure to Kanji, the Japanese language, or the Kanjiru interface.
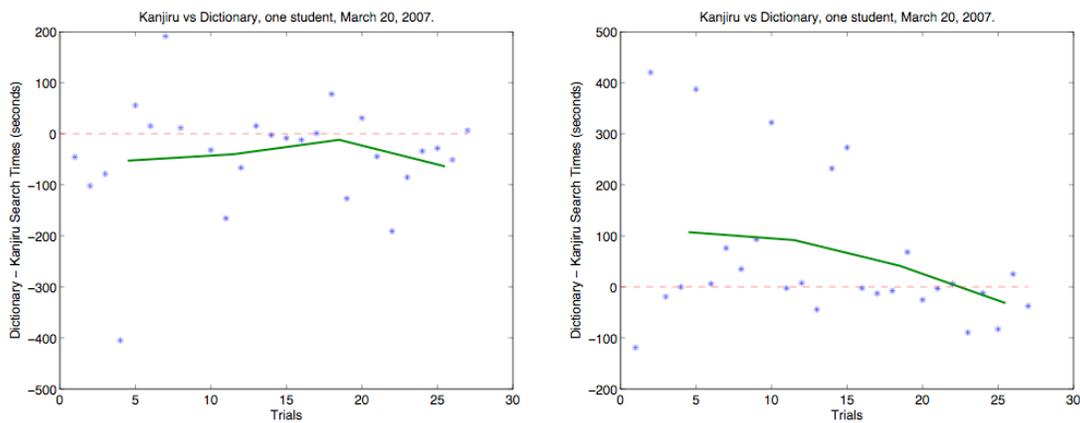

Figure 4: Comparison of dictionary and Kanjiru search times for two students with prior Kanji training. Negative times indicate that Kanjiru is faster than the dictionary.

REFERENCES

Breen, J. (2007). *Jim Breen's WWWJDIC Japanese-English Dictionary Server (Monash University Site)*, from www.csse.monash.edu.au/~jwb/wwwjdic.html.

Ward, D. J. and MacKay, D. J. (2002). *Fast hands-free writing by gaze direction. Nature*, 418, 838.

Kodansha (1991). *Kodansha's Compact Kanji Guide*. Kodansha International.