

Development and Evaluation of the WithAir Mobile Search Engine

Hideki Kawai

Internet Systems Research Laboratories, NEC Corp.
8916-47, Takayama-cho, Ikoma-city, Nara, Japan
+81-743-72-3684
h-kawai@ab.jp.nec.com

Susumu Akamine

Internet Systems Research Laboratories, NEC Corp.
8916-47, Takayama-cho, Ikoma-city, Nara, Japan
+81-743-72-3760
s-akamine@ak.jp.nec.com

Koji Kida

Internet Systems Research Laboratories, NEC Corp.
8916-47, Takayama-cho, Ikoma-city, Nara, Japan
+81-743-72-3677
kida@da.jp.nec.com

Katsushi Matsuda

Internet Systems Research Laboratories, NEC Corp.
8916-47, Takayama-cho, Ikoma-city, Nara, Japan
+81-743-72-3763
mat@bp.jp.nec.com

Toshikazu Fukushima

Internet Systems Research Laboratories, NEC Corp.
8916-47, Takayama-cho, Ikoma-city, Nara, Japan
+81-743-72-3756
t-fukushima@cj.jp.nec.com

ABSTRACT

This paper describes a novel search engine, WithAir that provides contents retrieval for i-mode, the most popular cell-phone service in Japan. WithAir allows users to complete typical search tasks with 4.4 keystrokes on average through keyword/purpose-predictive navigation. These two forms of navigation simplify, respectively, keyword input and search result output. WithAir has been used to search 1.3 million i-mode pages collected by document-type-based focused crawlers that can distinguish between i-mode pages and web pages with a precision of 99%. Additionally, WithAir offers fresh and exhaustive searching of 370,000 regional information pages extracted from 1.3 million i-mode pages with a precision of 93%.

Keywords

Mobile Search Engine, Input Support, Focused Crawler, Regional Information Search

1. INTRODUCTION

As the popularity of mobile devices such as cell-phones has grown, i-mode has become one of the most successful cell-phone services in the world. (Figure 1 shows the growth in the number of i-mode users and sites [1].) First introduced in Japan in February 1999 by NTT DoCoMo, i-mode now has more than 30 million subscribers. Currently, there are over 52,000 i-mode sites, and the i-mode traffic is over 300 million page views per day. This growth has led to demands for a better i-mode search engine. In this paper, we describe a mobile search engine, WithAir, developed based on four design concepts: (A) simplifying keyword input to reduce the cost of key operation, (B) simplifying

search result output to help users find relevant pages easily, (C) crawling i-mode pages automatically to constantly update and expand search results, and (D) offering regional information searching to help users find information on, for example, hotels or restaurants, while they are out.

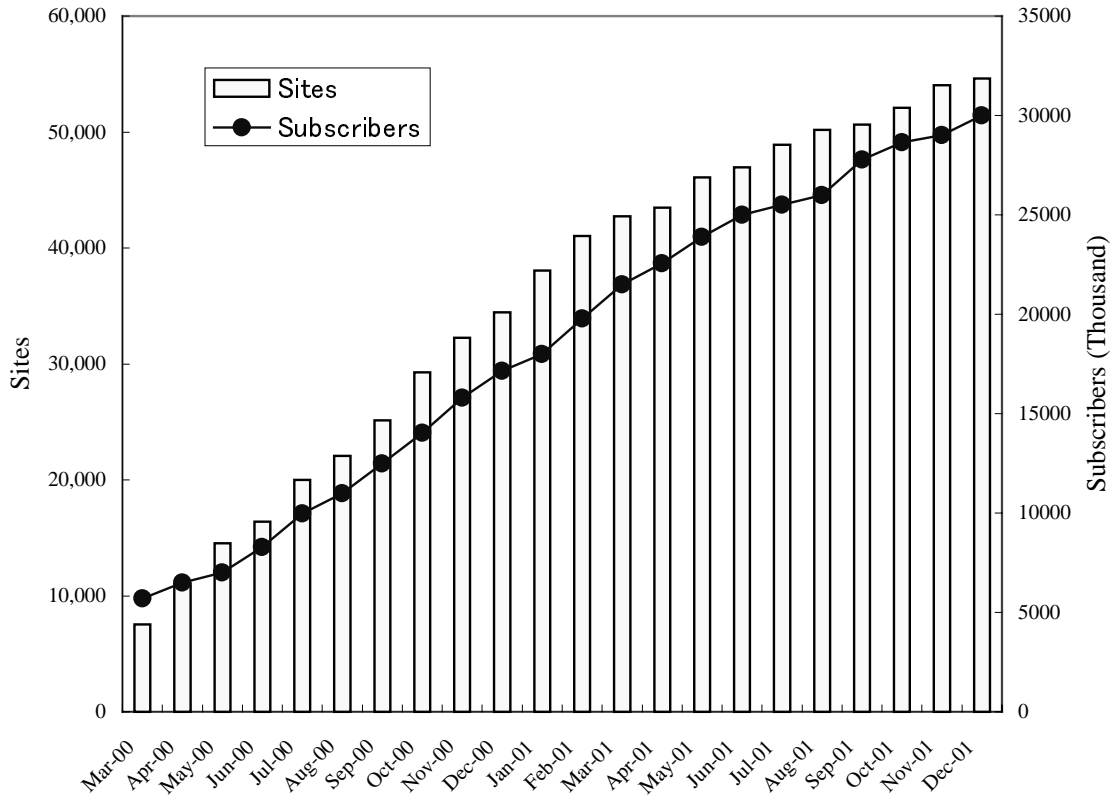


Figure 1: Increasing i-mode subscribers and sites [1]

2. SYSTEM OVERVIEW

WithAir has four important features based on the design concepts.

(1) Keyword predictive navigation reduces the operation cost of entering search keywords

The input method is a major issue for cell-phone device users. Inputting keywords consisting of alphabetical and Japanese characters with an alphanumeric keypad is very troublesome. The keyword predictive navigation of WithAir helps users input frequently used keywords taken from the query history.

(2) Purpose-predictive navigation reduces the trial-and-error inherent in browsing search results

The i-mode screen size does not permit the display of search results with detailed information like URLs or captions. To address this problem, the purpose-predictive navigation [6] of WithAir shows search results according to several typical search purposes of the user by anticipating from the input keyword. Since the search purposes can be expressed in one or two words, users can find

relevant pages at a glance. This alternative form of navigation guides users quickly and accurately on a small screen.

(3) Focused crawling i-mode pages based on document-type classification

I-mode pages are written in the Compact Hypertext Markup Language (CHTML) [2,3] which is a subset of the Hypertext Markup Language (HTML) used for typical web pages. The Hypertext Transfer Protocol (HTTP) is used to exchange i-mode page files as well as web pages. Therefore, general web crawlers tend to collect both i-mode and web pages. We used document-type classification [4] to selectively crawl i-mode pages because it can distinguish i-mode pages from web pages by scoring structural characteristics like CHTML tags, URLs, page size, and so on. For example, a page gains points if its URL contains /i/ or /i_mode/, while a page loses points if it contains tags unsupported by CHTML, such as <FRAME>. Higher scoring pages are more likely to be i-mode pages.

(4) Regional information classification of i-mode pages

People on the go often have a great need for regional information regarding hotels, restaurants, and so on. WithAir offers regional information searching by classifying i-mode pages according to the place names they contain, including addresses, zip codes, phone numbers, or train/subway station names [5]. A page that contains more place-names related to a particular region is considered more likely to contain information related to that region.

Figure 2 shows snapshots of the WithAir display for two sample scenarios that demonstrate the information discovery tasks that can be done using our system. (Though the output of WithAir is in Japanese, those snapshots are written in English for illustrative purposes.) One scenario is getting music information through a keyword search, and the other is getting travel information for Tokyo through a regional information search.

Imagine a user who has a general interest in music, and wants to enter a keyword "MUSIC" for the search. The user can use keyword predictive navigation (Feature (1)) by clicking the "KeywordNavi" link in the top page of WithAir (Fig. 2(a)). When character classes [MNO][TUV] are designated, the five highest frequency keywords that start with [MNO][TUV] are listed (e.g., "**NTT**", "**NTT DOCOMO**", "**MUSIC**", "**NTT EAST**" and "**NTV**" in Fig. 2(b)). The user can search with the keyword "MUSIC" by clicking the listed keyword "MUSIC". Purpose-predictive navigation (Feature (2)) then anticipates the intrinsic purpose of the user from statistical analysis of the keyword "MUSIC" and words that co-occur in the query history, and returns several sets of purpose and relevant sites related to the keyword "MUSIC" (e.g., in Fig. 2(c), "CD Shop" is purpose anticipated from the keyword "MUSIC" and "Music Navi" is a CD shop site). Below the purpose-predictive navigation results, there are results of a keyword search. Whole pages in the result are designed for i-mode, so the user can smoothly browse these pages (Feature (3)).

Our next scenario is a visitor to Tokyo who wants travel information. The user can use regional information search (Feature (4)) by clicking the "Regional Info" link in the top page (Fig. 2(a)). All the user has to do is to select a regional directory (Fig. 2(d)). WithAir then returns a selection of i-mode pages related to Tokyo (Fig. 2(e)). (Note that purpose predictive navigation is also available.) If the user is looking for a hotel to stay at, he would chose the second page in the Regional Info results which is titled "Tokyo <- HOTEL GUIDE".

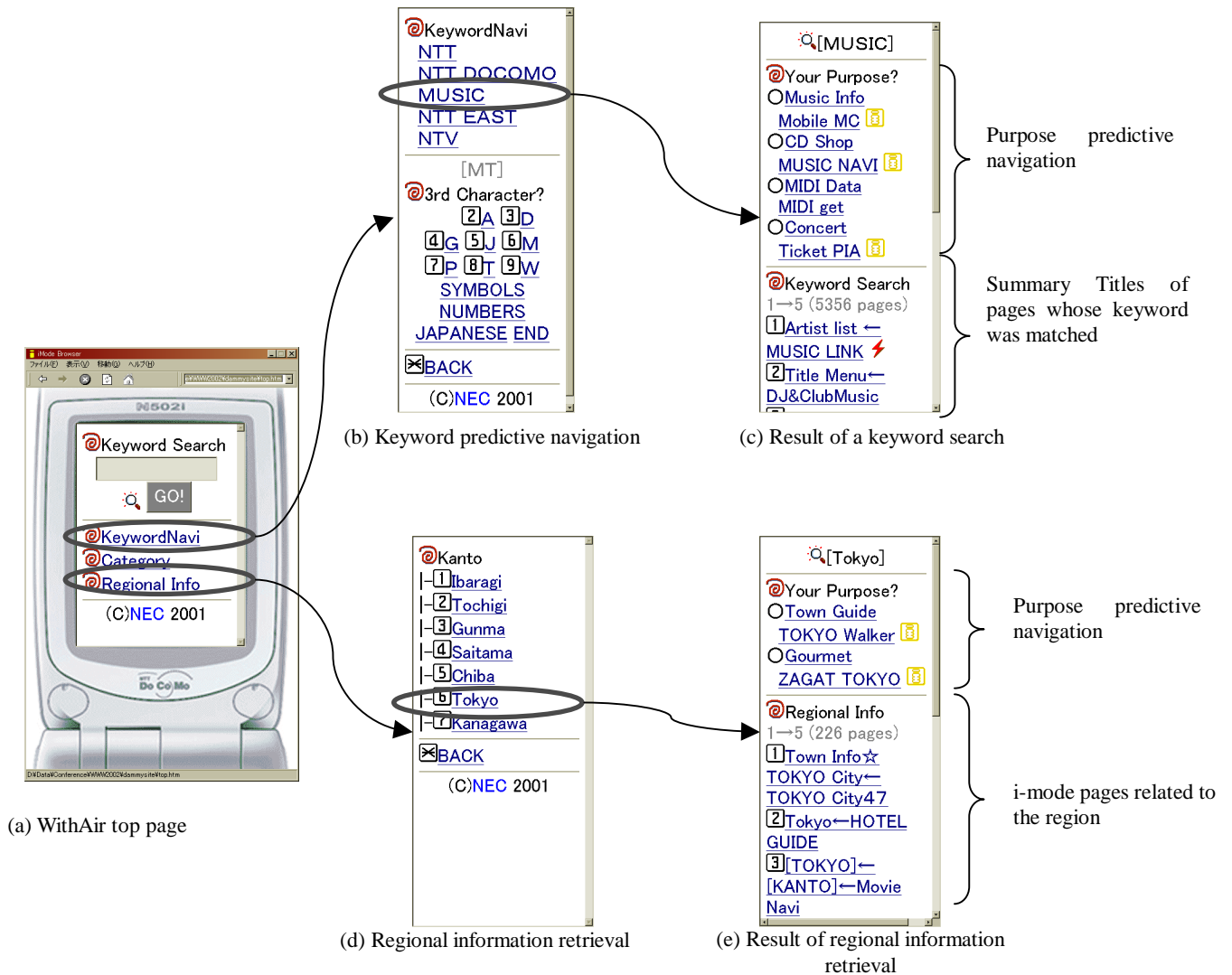


Figure 2: Screen-shots of WithAir. The user can use keyword predictive navigation by clicking the KeywordNavi link on the top page (a). If there is a keyword that the user wants to use from among the five highest frequency keywords on page (b), the user can search with that keyword by clicking it. When the user clicks the keyword MUSIC, WithAir displays purpose-predictive navigation and keyword search results for MUSIC as page (c). The user can also search for regional information by clicking the Regional Info link on page (a). After a region is selected on page (d), i-mode pages related to that region are listed as page (e).

3. EVALUATION

We measured the precision and recall of the i-mode focused crawler and regional information classification, and evaluated the user friendliness compared to conventional search engines.

3.1 Accuracy of focused crawler and regional information classification

Tables 1 and 2 show, respectively, the accuracy of the i-mode focused crawler and the regional information classification. About 1.3 million pages were collected and classified as i-mode pages by the crawler, and about 370,000 of these were classified as containing regional

information. The precision of the focused crawler was measured as the ratio of actual i-mode pages (1246 pages) to the total number of sampled pages classified as i-mode pages (1251 pages) by the crawler. Since we cannot know how many i-mode pages are actually existing in the web, we measured the recall of the focused crawler virtually as the ratio of pages classified as i-mode pages (1169 pages) to the total number of actual i-mode pages sampled (1326 pages). The precision and recall of regional information classification were calculated in the same way as those of the focused crawler.

Table 1: Performance of i-mode focused crawler

Gathered i-mode pages	Precision	Recall
1,300,000	99% (1246/1251)	88% (1169/1326)

Table 2: Performance of regional information extractor

Classified pages	Precision	Recall
370,000	93% (1006/1024)	76% (1480/1957)

3.2 User Friendliness

We counted the minimum keystrokes needed to accomplish a typical search task (e.g., "find a hotel room in Osaka", "find a bullet-train timetable") as the user friendliness index. Figure 3 compares our results for WithAir with those when using other i-mode search engines such as Yahoo! mobile [7] and Google for i-mode [8].

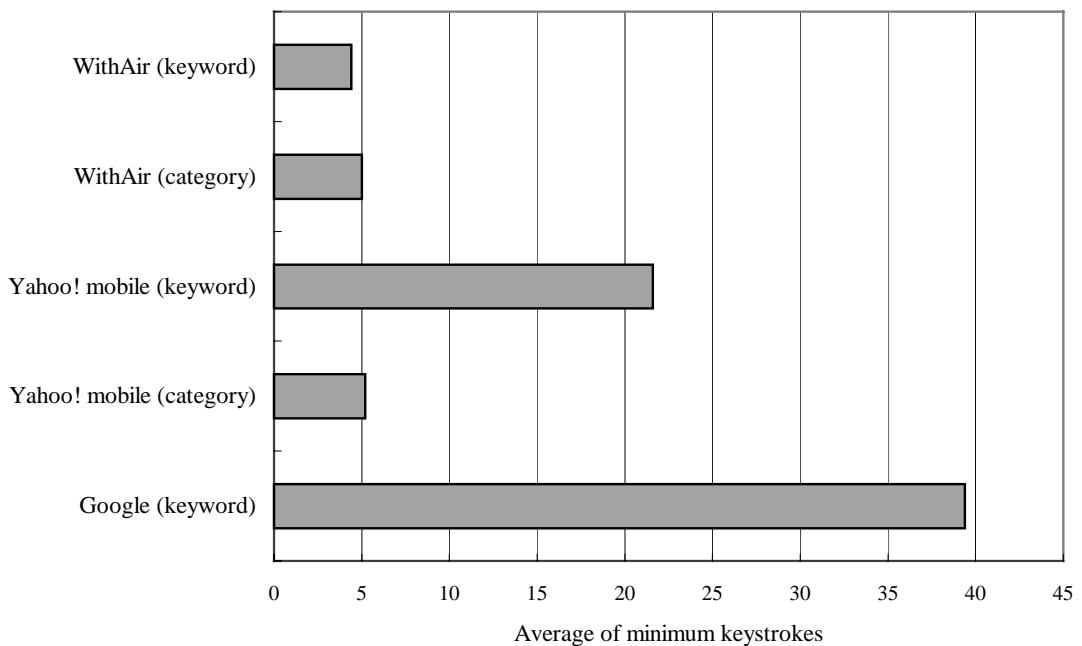


Figure 3: Comparison of minimum keystrokes

As shown, a keyword search on WithAir took from one-ninth to one-fifth as many keystrokes as with the other keyword search engines. This reduction is due to both the keyword predictive navigation, which reduces the keyword-entering cost, and the purpose-predictive navigation, which reduces the results-browsing cost. (Note, however, that users unfamiliar with the structure of the given category usually need to use more than the minimum keystrokes.) We therefore concluded that a keyword search on WithAir is currently the most user-friendly.

4. CONCLUSION

We have developed and implemented the WithAir i-mode search engine whose i-mode page focused crawler could collect 1.3 million i-mode pages with a precision of 99%, and whose regional information extractor could extract 370,000 regional information pages from i-mode pages with a precision of 93%. The keyword search on WithAir allows users to accomplish typical tasks with the fewest possible keystrokes.

WithAir was implemented in the i-mode search service of BIGLOBE [9], a major Internet service provider in Japan, on 24th July 2001. In the future, we will continue to evaluate the online performance of WithAir through actual test cases based on user feedback and access logs to improve the WithAir system and data processing.

REFERENCES

- [1] NTT DoCoMo: subscriber growth,
http://www.nttdocomo.com/html/subscriber_growth.html
- [2] NTT DoCoMo: i-mode service guideline,
<http://www.nttdocomo.com/html/disclaimer.html>
- [3] Kamada T, "Compact HTML for Small Information Appliances", W3C NOTE 09-Feb-1998,
<http://www.w3.org/TR/1998/NOTE-compactHTML-19980209/>
- [4] Matsuda K, Fukushima T, "Task-Oriented World Wide Web Retrieval by Document Type Classification", Proceedings of CIKM'99, 109-113 (1999).
- [5] Kida K, Ishiguro Y, Yamada H, Fukushima T, Matsuda K, "Development of the Mobile Location Search Engine", The 61st IPSJ National Conference, 1U-2, in Japanese (2000).
- [6] Kawai H, Akamine S, Kida K, Fukushima T, "Development of a mobile-oriented WWW search engine WithAir (2) -navigation-", The 62nd IPSJ National Conference, 6W-9, in Japanese (2001).
- [7] Yahoo! mobile, <http://mobile.yahoo.co.jp/>
- [8] Google for i-mode, <http://www.google.com/imode>
- [9] BIGLOBE Search Attayo for i-mode, <http://attayo.jp/i/>