Assisting Pictogram Selection with Semantic Interpretation

Heeryon Cho[1], Toru Ishida[1], Toshiyuki Takasaki[2], and Satoshi Oyama[1]

[1] Department of Social Informatics, Kyoto University, Kyoto 606-8501 Japan
cho@ai.soc.i.kyoto-u.ac.jp, {ishida,oyama}@i.kyoto-u.ac.jp
[2] Kyoto R&D Center, NPO Pangaea, Kyoto 600-8411 Japan
toshi@pangaean.org

Abstract. Participants at both end of the communication channel must share common pictogram interpretation to communicate. However, because pictogram interpretation can be ambiguous, pictogram communication can sometimes be difficult. To assist human task of selecting pictograms more likely to be interpreted as intended, we propose a *semantic relevance measure* which calculates how relevant a pictogram is to a given interpretation. The proposed measure uses pictogram interpretations and frequencies gathered from a web survey to define probability and similarity measurement of interpretation words. Moreover, the proposed measure is applied to categorized pictogram interpretations to enhance retrieval performance. Five pictogram categories are created using the five first level categories defined in the Concept Dictionary of EDR Electronic Dictionary. Retrieval performance among not-categorized interpretations, categorized and not-weighted interpretations, and categorized and weighted interpretations using semantic relevance measure were compared, and the categorized and weighted semantic relevance retrieval approach exhibited the highest F_1 measure and recall.

1 Introduction

Tags are prevalent form of metadata used in various applications today, describing, summarizing, or imparting additional meaning to the content to better assist content management by both humans and machines. Among various applications that incorporate tags, we focus on a pictogram email system which allows children to communicate to one another using pictogram messages[1,2]. Our goal is to support smooth pictogram communication between children, and to realize this, we focus on the pictogram selection stage where children select individual pictograms to create pictogram messages.

Pictogram is an icon which has a clear pictorial similarity with some object[3], and one who can recognize the object depicted in the pictogram can interpret the meaning associated with the object. Pictorial symbols, however, are not universally interpretable. A simple design like an arrow is often used to show direction, but there is no reason to believe that arrows suggest directionality to all people; they might also be taken as a symbol for war or bad luck[4].

S. Bechhofer et al.(Eds.): ESWC 2008, LNCS 5021, pp. 65–79, 2008.

Since the act of selecting a pictogram is done with a purpose of conveying certain meaning to the counterpart, the selected pictogram must carry intended meaning to both the sender and receiver of communication; that is, the selected pictogram must be relevant to the participants at both end of communication channel in order for the pictogram communication to be successful.

To assist pictogram selection, we propose a categorized usage of human-provided pictogram interpretations. Related research unifies the browsing by tags and visual features for intuitive exploration of image databases[5]. Our approach utilizes *categorized* pictogram interpretations together with the *semantic relevance measure* to retrieve and rank relevant pictograms for a given interpretation. We define *pictogram categories* by appropriating first level categories defined in the Concept Dictionary of EDR Electronic Dictionary[6]. We will show that categorized and weighted semantic relevance approach returns better result than not-categorized, not-weighted approaches.

In the following section, five pictogram categories are described, and characteristics in pictogram interpretation are clarified. Section 3 describes semantic relevance measure, and categorization and weighting of interpretation words. Section 4 presents precision, recall, and retrieval examples of four pictogram retrieval approaches. A prototype implementation is also presented. Finally, section 5 concludes this paper.

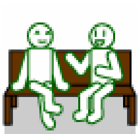
2 Ambiguity in Pictogram Interpretation

A twenty-five month pictogram web survey was conducted from October 1st, 2005 to November 7th, 2007 to collect free-answer English pictogram interpretation words or phrases from respondents living in the United States. Tallying the unique username–IP address pairs, a total of 1,602 respondents participated in the survey. Details of the earlier survey can be found in [7,8].

2.1 Polysemous Interpretation

From the pictogram web survey data, interpretations consisting of English words or phrases were tallied according to unique interpretation words. Phrasal expressions and misspellings were discarded. An example of tallied pictogram interpretation words is shown in Table 1. As shown, a pictogram can have multiple interpretations which include both similar and different-meaning interpretations. For example, words like *talking*, *talk*, *chatting*, *conversing*, *chat*, and *communicating* are all similar action-related words describing the act of speaking. Other action-related words are *date*, *flirt*, *sit*, *flirting*, *listening*, *love*, and *play*. On the other hand, when the focus shifts to the people depicted in the pictogram, the pictogram is interpreted as *friends* or *family*. Or it can be interpreted as some kind of place such as *church*, or as an emotional state such as *happy*. One way to organize mixed interpretations containing both similar and different meanings is to group them into categories. We use the Concept Dictionary in the EDR Electronic Dictionary[6] to group interpretation words into five first level categories

Table 1. An example of tallied pictogram interpretation words

PICTOGRAM	WORD	FREQ	RATIO	WORD	FREQ	RATIO
	talking	58	0.367	church	1	0.006
	talk	27	0.171	communication	1	0.006
	conversation	20	0.127	family	1	0.006
	friends	15	0.095	flirting	1	0.006
	chatting	13	0.082	friend	1	0.006
	conversing	5	0.032	friendly	1	0.006
	chat	2	0.013	happy	1	0.006
	communicating	2	0.013	listening	1	0.006
	date	2	0.013	love	1	0.006
	flirt	2	0.013	play	1	0.006
	sit	2	0.013			
	TOTAL FREQUENCY = 158, TOTAL RATIO = 0.999					

defined in the dictionary. We borrow these five first level categories to define five *pictogram categories*.

The EDR Electronic Dictionary was developed for advanced processing of natural language by computers, and is composed of five types of dictionaries (Word, Bilingual, Concept, Co-occurrence, and Technical Terminology), as well as the EDR Corpus. The Concept Dictionary contains information on the approximately 410,000 concepts listed in the Word Dictionary and is divided according to information type into the Headconcept Dictionary, the Concept Classification Dictionary, and the Concept Description Dictionary. The Headconcept Dictionary describes information on the concepts themselves. The Concept Classification Dictionary describes the super-sub relations among the approximately 410,000 concepts. The “super-sub” relation refers to the inclusion relation between concepts, and the set of interlinked concepts can be regarded as a type of thesaurus. The Concept Description Dictionary describes the semantic (binary) relations, such as ‘agent’, ‘implement’, and ‘place’, between concepts that co-occur in a sentence[6]. We use the Headconcept Dictionary and the Concept Classification Dictionary to obtain super-concepts of the pictogram interpretation words. A record in the Headconcept Dictionary is composed of a record number, a concept identifier, an English headconcept, a Japanese headconcept, an English concept explication, a Japanese concept explication, and a management information[1]. Below shows two records containing the English headconcept “talk”[2]. Notice that there are two different concept identifiers (0dc0d6, 0dc0d7) for the same English headconcept “talk”[3]:

- `CPH0144055 0dc0d6 talk JH “an informal speech” “JE” DATE=“95/6/6”`
- `CPH0144056 0dc0d7 talk JH “a topic for discussion” “JE” DATE=“95/6/6”`

[1] A headconcept is a word whose meaning most closely expresses the meaning of the concept. A concept explication is an explanation of the concept’s meaning.

[2] `JH` indicates Japanese headconcept and `"JE"` indicates Japanese explication.

[3] Overall, there are 13 concept identifiers matching the English headconcept “talk”.

We obtain concept identifier(s) of a pictogram interpretation word by matching the interpretation word string to English headconcept string in the Headconcept Dictionary. Once the concept identifier(s) are obtained, we use the Concept Classification Dictionary to retrieve the first level categories of the concept identifier(s). A record in the Concept Classification Dictionary is composed of a record number, a concept identifier of the super-concept, a concept identifier of the sub-concept, and management information. Below shows two concept classification dictionary records containing the super-sub concept identifiers[4]:

- `CPC0144500 444059 0dc0d6 DATE="95/6/7"`
- `CPC0419183 443e79 444059 DATE="95/6/7"`

Note that there may be more than one super-concept (identifier) for a given concept (identifier) since the EDR Concept Dictionary allows multiple inheritance. By climbing up the super-concept of a given concept, we reach the root concept which is 3aa966 'concept'. Five categories defined at the first level, placed just below the root concept, will be used as five pictogram categories for categorizing pictogram interpretation words. The headings of the five categories are:

(a) Human or subject whose behavior (actions) resembles that of a human
(b) Matter or an affair
(c) Event/occurrence
(d) Location/locale/place
(e) Time

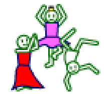
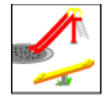
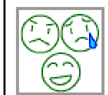
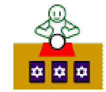
Superclasses in SUMO ontology[9] could be another candidate for defining pictogram categories, but we chose EDR because we needed to handle both English and Japanese pictogram interpretations. For brevity, we abbreviate the pictogram category headings as (a) AGENT, (b) MATTER, (c) EVENT, (d) LOCATION, and (e) TIME. Table 2 shows examples of nine pictograms' interpretation words categorized into the five pictogram categories. Each column contains interpretation words of each pictogram. Each cell on the same row contains interpretation words for each of the five categories. One interpretation word can be assigned to multiple categories, but in the case of Table 2, each word is assigned to a single *major category*. Major category is explained later in section 3.2. We now look at each category in detail.

(a) AGENT. Pictograms containing human figures can trigger interpretations explaining something about a person or people. Table 2 AGENT row contains words like *family*, *dancers*, *people*, *crowd*, *fortune teller*, and *magician* which all explain a specific kind of person or people.

(b) MATTER. Concrete objects or objective subjects are indicated. Table 2 MATTER row contains words like *good morning*, *good night*, *moon*, *good evening*, *dancing*, *chicken*, *picture*, *ballet*, *card*, *drama*, *crystal ball*, and *magic* which point to some physical object(s) or subject depicted in the pictogram.

[4] 444059 is the super-concept identifier of 0dc0d6, and 443e79 is the super-concept identifier of 444059.

Table 2. Polysemous interpretations (each column) and shared interpretations (boldface type)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
									
AGENT					**family**	**family** dancers		people **family** crowd	fortune teller magician
MATTER	good morning **good night**	**moon** good evening **dancing** **good night**	**moon** **good night**	chicken **picture**		**dancing** ballet		**card** drama **picture**	**card** crystal ball magic
EVENT	**talking** **sleeping** **happy** **play** **friendly**	**talking** **happy** **play**	**sleeping** dream peaceful	**sleeping**	**talking** date **play** **friendly**	dance jumping **play**	slide fun **play**	crying **happy** mixed	bowling guess **play**
LOCATION							playground	theater	
TIME	**morning** day **bedtime**	**night** **evening**	**night** **bedtime**	**morning** **night** **evening**					future

(c) EVENT. Actions or states are captured and described. Table 2 EVENT row contains words like *talking*, *sleeping*, *happy*, *play*, *friendly*, *dream*, *peaceful*, *date*, *dance*, *jumping*, *slide*, *fun*, *crying*, *mixed*, *bowling*, and *guess* which all convey present states or ongoing actions.

(d) LOCATION. Place, setting, or background of the pictogram are on focus rather than the object occupying the center or the foreground of the setting. Table 2 LOCATION row contains words like *playground* and *theater* which all indicate specific places or settings relevant to the central object(s).

(e) TIME. Time-related concepts are sometimes perceived. Table 2 TIME row contains words like *morning*, *day*, *bedtime*, *night*, *evening*, and *future* which all convey specific moments in time.

Categorizing the words into five pictogram categories elucidates two key aspects of polysemy in pictogram interpretation. Firstly, interpretations spread across different categories lead to different meanings. For example, interpretation words in Table 2 column (8) include AGENT category's *crowd* which describes a specific relationship between people, EVENT category's *crying* which describes an ongoing action, and LOCATION category's *theater* which describes a place for presenting a show, and they all mean very different things. This is due to the different focus of attention given by each individual.

Secondly, while interpretation words placed within the same category may contain similar words such as *sleeping* and *dream* (Table 2 column (3) row EVENT), or *dance* and *jumping* (Table 2 column (6) row EVENT), contrasting or opposite-meaning words sometimes coexist within the same category. For example, Table 2 column (4) row TIME contains both *morning* and *night*, which are contrasting time-related concepts, and column (1) row MATTER contains both *good morning* and *good night*, which are contrasting greeting words.

While the words in Table 2 column (2) row TIME are varied yet similar (*night* and *evening*), the words in column (4) row TIME are confusing because contrasting interpretations are given on the same viewpoint (*morning* and *night*). To summarize the above findings, it can be said that polysemy in pictogram interpretation is dependent on the interpreter's perspective; usually, interpretations differ across different categories or perspectives, but sometimes interpretations may vary even within the same category or perspective.

When a pictogram having polysemous interpretations is used in communication, there is a possibility that a sender and receiver might interpret the same pictogram differently. In the case of pictogram in Table 2 column (4), it could be interpreted quite differently as *morning* and *night* by the sender and receiver. One way to assist the sender to choose a pictogram with higher chance of conveying the intended message to the receiver is to display possible interpretations of a given pictogram. If various possible interpretations are presented, the sender can speculate receiver's interpretation before choosing and using the pictogram. For example, if the sender knows a priori that Table 2 pictogram (4) can be interpreted as both *morning* and *night*, s/he can guess ahead that it might be interpreted oppositely by the receiver, and avoid choosing the pictogram.

We will refer to this characteristic of one-to-many correspondence in pictogram-to-meaning and an associative measure of displaying possible pictogram interpretations as assisting selection of pictograms having polysemous interpretations.

2.2 Shared Interpretation

One pictogram may have various interpretations, but these interpretations are not necessarily different across different pictograms. Sometimes multiple pictograms share common interpretation(s) among themselves. Words indicated in boldface type in Table 2 are such interpretations shared by more than one pictogram: Table 2 (5), (6), and (8) share *family* (row AGENT); (1), (2), and (3) share *good night* (row MATTER); (1) and (5) share *friendly* (row EVENT); and (2), (3), and (4) share *night* (row TIME) and so forth.

The fact that multiple pictograms can share common interpretation implies that each one of these pictograms can be interpreted as such. The degree to which each is interpreted, however, may vary according to the pictogram. For example, Table 2 pictograms (1), (2), and (3) can all be interpreted as *good night* (row MATTER), but (1) can also be interpreted as *good morning* while (2) can also be interpreted as *good evening*. Furthermore, if we move down the table to row TIME, we see that (1) has *morning* as time-related interpretation while (2) and (3) have *night*. Suppose two people A and B each use pictogram (1) and (2) respectively to send a "good night" message to person C. Upon receiving the message, however, C may interpret the two messages as "good morning" for A and "good night" for B. Even though A and B both intend on conveying a "good night" message, it may not always be the case that C will interpret the two pictograms likewise. This is because the degree of interpretation may vary across similar-meaning pictograms; one reason may be due to other possible interpretations within the pictogram (as in *good morning* and *good evening*).

One way to assist the selection of pictograms among multiple similar-meaning pictograms is to rank those pictograms according to the degree of relevancy of a pictogram to a given interpretation. Presenting ranked pictograms to the user who selects the pictogram to be used in communication will allow the user to understand which pictogram is most likely to be interpreted as intended. In order to rank pictograms according to the interpretation relevancy, some kind of metric which measures the relevancy of a pictogram to an interpretation is needed. We will refer to this characteristic of one-to-many correspondence in meaning-to-pictogram and an associative measure of ranking pictograms according to interpretation relevancy as assisting selection of pictograms having shared interpretations.

3 Semantic Relevance Measure

We identified ambiguities in pictogram interpretation and possible issues involved in the usage of such pictograms in communication. Here, we propose a *semantic relevance measure* which outputs relevancy values of each pictogram when a pictogram interpretation is given. Our method presupposes a set of pictograms having a list of interpretation words and ratios for each pictogram.

3.1 Definition

We assume that pictograms each have a list of interpretation words and frequencies as the one shown in Table 1. Each unique interpretation word has a frequency. Each word frequency indicates the number of people who answered the pictogram to have that interpretation. The ratio of an interpretation word, which can be calculated by dividing the word frequency by the total word frequency of that pictogram, indicates how much support people give to that interpretation. For example, in the case of pictogram in Table 1, it can be said that more people support *talking* (58 out of 158) as the interpretation for the given pictogram than *sit* (2 out of 158). The higher the ratio of a specific interpretation word of a pictogram, the more that pictogram is accepted by people for that interpretation.

We define semantic relevance of a pictogram to be the measure of relevancy between a word query and interpretation words of a pictogram. Let $w_1, w_2, ..., w_n$ be interpretation words of pictogram e. Let the ratio of each interpretation word in a pictogram to be $P(w_1|e), P(w_2|e), ..., P(w_n|e)$. For example, the ratio of the interpretation word *talking* for the pictogram in Table 1 can be calculated as $P(talking|e) = 58/158$. Then the simplest equation that assesses the relevancy of a pictogram e in relation to a query w_i can be defined as follows:

$$P(w_i|e) \tag{1}$$

This equation, however, does not take into account the similarity of interpretation words. For instance, when "talking" is given as query, pictograms having similar interpretation word like "gossiping", but not "talking" fail to be measured as relevant when only the ratio is considered. To solve this, we need to define $similarity(w_i, w_j)$ between interpretation words in some way. Using the similarity, we can define the measure of *Semantic Relevance* or $SR(w_i, e)$ as follows:

$$SR(w_i, e) = \sum_j P(w_j|e)similarity(w_i, w_j) \tag{2}$$

There are several similarity measures. We draw upon the definition of similarity given by Lin[10] which states that similarity between A and B is measured by the ratio between the information needed to state the commonality of A and B and the information needed to fully describe what A and B are. Here, we calculate the similarity of w_i and w_j by figuring out how many pictograms contain certain interpretation words. When there is a pictogram set E_i having an interpretation word w_i, the similarity between interpretation word w_i and w_j can be defined as follows:

$$similarity(w_i, w_j) = |E_i \cap E_j|/|E_i \cup E_j| \tag{3}$$

$|E_i \cap E_j|$ is the number of pictograms having both w_i and w_j as interpretation words. $|E_i \cup E_j|$ is the number of pictograms having either w_i or w_j as interpretation words. Based on (2) and (3), the semantic relevance or the measure of relevancy to return pictogram e when w_i is input as query can be calculated as follows:

$$SR(w_i, e) = \sum_j P(w_j|e)|E_i \cap E_j|/|E_i \cup E_j| \tag{4}$$

The resulting semantic relevance values will fall between one and zero, which means either a pictogram is completely relevant to the interpretation or completely irrelevant. Using the semantic relevance values, pictograms can be ranked from very relevant (value close to 1) to not so relevant (value close to 0). As the value nears zero, the pictogram becomes less relevant; hence, a cutoff point is needed to discard the less relevant pictograms. Setting an ideal cutoff point that satisfies all interpretations and pictograms is difficult, however, since all words contained in a pictogram, regardless of relation to each other, each influence the calculation. For example, let's say that we want to find a pictogram which can convey the meaning "friend" or "friends". Pictogram in Table 1 could be a candidate since it contains both words with a total ratio of 0.1. When the semantic relevance is calculated, however, the equation takes into account all the interpretation words including *talking* or *church* or *play*. Selecting a set of words relevant to the query would reduce the effect of less-relevant interpretation words affecting the calculation. Based on this prediction, we propose a semantic relevance calculation on categorized interpretations.

3.2 Word Categorization, Word Weighting, and Result Ranking

Word Categorization. Interpretation words are categorized into five pictogram categories described in section 2.1. Note that some headconcept(s) in the EDR Electronic Dictionary link to multiple concepts, and some concepts lead to multiple super-concepts (i.e. multiple inheritance). For example, in the case of the word (headconcept) *park*, three kinds of pictogram categories are obtained repeatedly: LOCATION category six times, MATTER category five times, and EVENT category four times. In such cases of multiple categories, we use all categories since we cannot accurately guess on the single correct category intended by each respondent who participated in the web survey.

Word Weighting. Although we cannot correctly decide on the single, intended category of a word, we can calculate the ratio of the pictogram category of each word. For example, in the case of *park*, the LOCATION category has the most number of repeated categories (six). Next is the MATTER category (five) followed by the EVENT category (four). In the case of the word *night*, the TIME category has most number of categories (seven) followed by EVENT (five) and MATTER (one). We can utilize such category constitution by calculating the ratio of the repeated categories and assigning the ratio as weights to the word in a given category. For example, the word *park* can be assigned to LOCATION, MATTER and EVENT category, and for each category, weights of 6/15, 5/15 and 4/15 can be assigned to the word. Same with *night*. The word *night* in the

TIME category will have the largest weight of 7/13 compared to EVENT (5/13) or MATTER (1/13). Consequently, the *major category* of *park* and *night* will be LOCATION and TIME respectively.

Result Ranking. Applying the semantic relevance calculation to categorized interpretations will return five semantic relevance values for each pictogram. We compare the highest value with the cutoff value to determine whether the pictogram is relevant or not. Once the relevant pictograms are selected, pictograms are then ranked according to the semantic relevance value of the query's major category. For example, if the query is "night", relevant pictograms are first selected using the highest semantic relevance value in each pictogram, and once candidate pictograms are selected, the pictograms are then ranked according to the semantic relevance value of the query's major category, which in this case is the TIME category. We use 0.5 cutoff value for the evaluation and prototype implementation described next.

4 Evaluation

Using the semantic relevance measure, retrieval tasks were performed to evaluate the semantic relevance measure and the categorized and weighted pictogram retrieval approach. Baseline for comparison was a simple string match of the query to interpretation words having a ratio greater than 0.5[5]. We also implemented a prototype web-based pictogram retrieval system (Fig. 1).

Comparison of Four Approaches. Four pictogram retrieval approaches were evaluated: (1) baseline approach which returns pictograms containing the query as interpretation word with ratio greater than 0.5; (2) semantic relevance approach which calculates semantic relevance value using not-categorized interpretations; (3) semantic relevance approach which calculates semantic relevance values using categorized interpretations; and (4) semantic relevance approach which calculates semantic relevance values using categorized and weighted interpretations. We wanted to see if (a) the fourth approach, the categorized and weighted approach, performed better than the rest; (b) the semantic relevance approach in general was better than the simple query match approach; (c) the categorized approach in general was better than the not-categorized approach.

Creation of Relevant Pictogram Set. A relevant pictogram set was created by five human judges who were all undergraduate students. There were 903 unique words for 120 pictograms, which meant that these words could be used as queries in the retrieval tasks. We performed retrieval tasks with these 903 words using the four approaches to filter out words that returned the same result among the four approaches, since those words would be ineffective in discerning the performance difference of the four approaches. A total of 399 words returned the same results for all four approaches. Another 216 words returned the same

[5] This is the same as selecting pictograms with $P(w_j|e) > 0.5$ where w_j equals query.

results for the three semantic relevance approaches. That left us with 288 words. Among the 288 words, words having more than 10 candidate pictograms, similar words (e.g. *hen*, *rooster*), singular/plural words (e.g. *girl*, *girls*), and varied tenses (e.g. *win*, *winning*) were filtered leaving 193 words to be judged for relevancy. For each of the 193 words, all pictograms containing the word were listed as candidate pictograms to be judged for relevancy.

A questionnaire containing each of the 193 words and candidate pictograms with ranked list of interpretation words[6] were given to five human judges, and they were to first judge whether each candidate pictogram can be interpreted as the given word, and then if judged relevant, write down the ranking among the relevant pictograms. The five judgments were analyzed by tallying the relevance judgments, and pictogram ranking was determined by calculating the averages and variances of the judgments[7]. After the five human judges' relevance judgments were analyzed, 30 words were additionally deleted since none of the candidate pictograms were judged as relevant. As a result, a ranked relevant pictogram set for 163 words was created and used in the evaluation[8].

Precision and Recall. Table 3 shows precision, recall, and F_1 measure of the four pictogram retrieval approaches. Each value is the mean performance value of 163 retrieval tasks performed[9]. A cutoff value of 0.5 was used for the three semantic relevance approaches. Based on the performance values listed in Table 3, we see that (a) the categorized and weighted semantic relevance approach performs better than the rest in terms of recall (0.70472) and F_1 measure (0.73757); (b) the semantic relevance approach in general performs much better than the simple query string match approach; and that (c) the categorized approach in general performs much better than the not-categorized approach.

It should be noted that the greatest gain in performance is achieved through the categorization of the interpretation words. By contrast, only a minimal gain is obtained through word-weighting as exhibited by the not-weighted vs. weighted performance values (e.g. 0.71492 vs. 0.73757 F_1 measure values). Through this

[6] The probability of each pictogram interpretation word was not displayed in the questionnaire, but was used to list the words with greater probability at the top.

[7] If three or more people judged relevant, the pictogram was judged relevant. Otherwise, the pictogram was discarded. Average ranking for each of the relevant pictogram was calculated. If average rankings were the same among multiple pictograms, variance was calculated and compared. The smaller the variance, the higher was the ranking.

[8] The composition ratio of the major category in 903, 288, 193 and 163 words were:
- 903 words: [AGENT, 9%], [MATTER,27%], [EVENT,56%], [LOCATION, 4%], [TIME,3%]
- 288 words: [AGENT,10%], [MATTER,26%], [EVENT,49%], [LOCATION, 8%], [TIME,6%]
- 193 words: [AGENT, 9%], [MATTER,28%], [EVENT,47%], [LOCATION,10%], [TIME,6%]
- 163 words: [AGENT, 9%], [MATTER,29%], [EVENT,47%], [LOCATION,10%], [TIME,6%]

[9] Mean precision value was calculated using the valid tasks that returned at least one result. The number of valid tasks for each approach was: QUERY MATCH = 9 tasks, *SR* WITHOUT CATEGORY = 49 tasks, *SR* WITH CATEGORY & NOT WEIGHTED = 139 tasks, and *SR* WITH CATEGORY & WEIGHTED = 153 tasks.

Table 3. Precision, recall, and F_1 measure of four pictogram retrieval approaches

APPROACH / MEASURE (MEAN FOR 163 TASKS)	QUERY MATCH RATIO>0.5	SEMANTIC RELEVANCE		
		WITHOUT CATEGORY	WITH CATEGORY	
			NOT-WEIGHTED	WEIGHTED
PRECISION	1.00000	0.87755	0.79808	0.77364
RECALL	0.02853	0.18108	0.64746	0.70472
F_1 MEASURE	0.05547	0.30022	0.71492	0.73757

example, we confirmed that a simple classification or pre-categorization of words can contribute greatly to the improvement of retrieval performance.

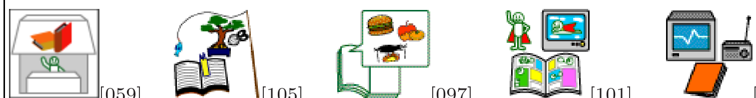
Examples of Retrieved Results. Table 4 shows pictogram retrieval results of five queries, "doctor", "book", "cry", "playground", and "bedtime", on four different approaches: (1) ALL PICTOGRAMS WITH QUERY lists all pictograms containing the query as interpretation word. The pictograms are sorted from the largest interpretation ratio to the smallest; (2) HUMAN JUDGED AS RELEVENT lists relevant pictograms selected by five human judges upon seeing the candidate pictograms listed in (1). The pictograms are ranked with the most relevant pictogram starting from the left. The pictograms listed here are the relevant pictogram set of the given word; (3) QUERY MATCH RATIO > 0.5 lists all pictograms having the query as interpretation word with ratio greater than 0.5; (4) *SR* WITHOUT CATEGORY uses not-categorized interpretations to calculate the semantic relevance value; (5) *SR* WITH CATEGORY & NOT-WEIGHTED uses categorized interpretations to calculate five semantic relevance values for each pictogram; (6) *SR* WITH CATEGORY & WEIGHTED uses categorized and weighted interpretations to calculate five semantic relevance values for each pictogram. In the three semantic relevance approaches (4), (5), and (6), a cutoff value of 0.5 was used. Once the semantic relevance values were calculated, the pictograms were ranked according to the semantic relevance value of the major category. Images of the candidate pictograms that contain query as interpretation word are listed at the bottom five rows of Table 4.

For instance, for the query "book" (Table 4 third column), there are two relevant pictograms [059, 097] out of five, and we see that only one approach, *SR* WITH CATEGORY & WEIGHTED, succeeds in returning the first relevant pictogram [059]. Similar readings can be done on the remaining four queries.

Prototype Implementation. Figure 1 shows a screenshot of a web-based pictogram retrieval system which uses the interpretation words collected from the survey. A search result for query "night" using the categorized and weighted approach is displayed. Contrasting interpretations, such as *night* and *morning* at the bottom right, are elucidated once the interpretations are categorized.

Table 4. Comparison of four approachs' retrieved results for five queries (Note: *SR* is an abbreviation for *Semantic Relevance*)

APPROACH \ QUERY	DOCTOR	BOOK	CRY	PLAYGROUND	BEDTIME
ALL PICTOGRAMS WITH QUERY	[043,044,090]	[059,105,097,101,102]	[007,006,050,073,072]	[039,113,118,105,038]	[003,001,009]
HUMAN JUDGED AS RELEVANT	[043,044]	[059,097]	[007,006]	[113,039,118,112]	[003,009]
QUERY MATCH RATIO > 0.5	[043]	[ ]	[ ]	[ ]	[ ]
SR WITHOUT CATEGORY	[043]	[ ]	[007]	[039]	[ ]
SR WITH CATEGORY & NOT-WEIGHTED	[043,044]	[ ]	[007,006]	[039,113]	[ ]
SR WITH CATEGORY & WEIGHTED	[043,044]	[059]	[007,006]	[039,118,113]	[003]

All pictograms containing DOCTOR	 [043], [044], [090]
All pictograms containing BOOK	 [059], [105], [097], [101], [102]
All pictograms containing CRY	 [007], [006], [050], [073], [072]
All pictograms containing PLAYGROUND	 [039], [113], [118], [105], [038]
All pictograms containing BEDTIME	 [003], [001], [009]

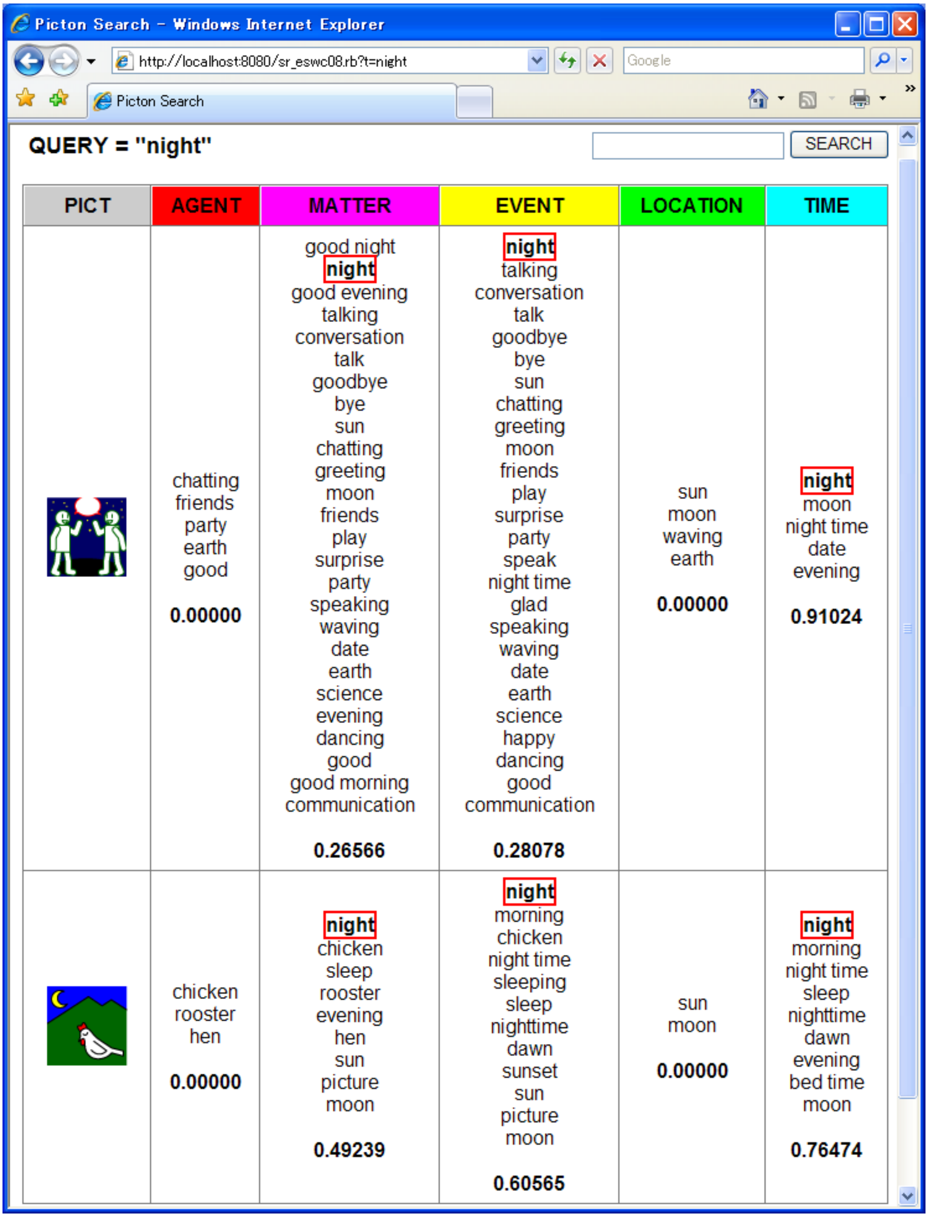


Fig. 1. A screenshot of web-based pictogram retrieval system prototype which uses the categorized and weighted semantic relevance approach with a 0.5 cutoff value. Results for the query "night" is displayed. Notice that contrasting interpretations, *night* and *morning* at the bottom right, become evident once the interpretations are categorized.

5 Conclusion

Pictograms used in a pictogram email system are created by novices at pictogram design, and they do not have single, clear semantics. To retrieve better intention-conveying pictograms using a word query, we proposed a semantic relevance measure which utilizes interpretation words and frequencies collected from a web survey. The proposed measure takes into account the probability and similarity in a set of pictogram interpretation words, and to enhance retrieval performance, pictogram interpretations were categorized into five pictogram categories using the Concept Dictionary in EDR Electronic Dictionary. The retrieval performance of (1) not-categorized, (2) categorized, and (3) categorized and weighted semantic relevance retrieval approaches were compared, and the categorized and weighted semantic relevance retrieval approach performed better than the rest.

Acknowledgements. This research was partially supported by the International Communications Foundation and the Global COE Program "Informatics Education and Research Center for Knowledge-Circulating Society."

References

1. Takasaki, T.: PictNet: Semantic infrastructure for pictogram communication. In: Sojka, P., Choi, K.S., Fellbaum, C., Vossen, P. (eds.) GWC 2006. Proc. 3rd Int'l WordNet Conf., pp. 279–284 (2006)
2. Takasaki, T.: Design and development of a pictogram communication system for children around the world. In: Ishida, T., R. Fussell, S., T. J. M. Vossen, P. (eds.) IWIC 2007. LNCS, vol. 4568, pp. 193–206. Springer, Heidelberg (2007)
3. Marcus, A.: Icons, symbols, and signs: Visible languages to facilitate communication. Interactions 10(3), 37–43 (2003)
4. Kolers, P.A.: Some formal characteristics of pictograms. American Scientist 57, 348–363 (1969)
5. Aurnhammer, M., Hanappe, P., Steels, L.: Augmenting navigation for collaborative tagging with emergent semantics. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 58–71. Springer, Heidelberg (2006)
6. National Institute of Information and Communications Technology (NICT): EDR Electronic Dictionary Version 2.0 Technical Guide
7. Cho, H., Ishida, T., Inaba, R., Takasaki, T., Mori, Y.: Pictogram retrieval based on collective semantics. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4552, pp. 31–39. Springer, Heidelberg (2007)
8. Cho, H., Ishida, T., Yamashita, N., Inaba, R., Mori, Y., Koda, T.: Culturally-situated pictogram retrieval. In: Ishida, T., R. Fussell, S., T. J. M. Vossen, P. (eds.) IWIC 2007. LNCS, vol. 4568, pp. 221–235. Springer, Heidelberg (2007)
9. Niles, I., Pease, A.: Towards a standard upper ontology. In: FOIS 2001. Proc. 2nd Int'l Conf. on Formal Ontology in Information Systems (2001)
10. Lin, D.: An information-theoretic definition of similarity. In: ICML 1998. Proc. of the 15th Int'l Conf. on Machine Learning, pp. 296–304 (1998)