Enriching Dictionaries with Images from the Internet - Targeting Wikipedia and a Japanese Semantic Lexicon: Lexeed -

Sanae Fujita
NTT Communication Science Lab.
`sanae@cslab.kecl.ntt.co.jp`

Masaaki Nagata
NTT Communication Science Lab.
`nagata.masaaki@lab.ntt.co.jp`

Abstract

We propose a simple but effective method for enriching dictionary definitions with images based on image searches. Various query expansion methods using synonyms/hypernyms (or related words) are evaluated. We demonstrate that our method is effective in obtaining high-precision images that complement dictionary entries, even for words with abstract or multiple meanings.

1 Introduction

The Internet is an immense resource for images. If we can form connections between these images and dictionary definitions, we can create rich dictionary resources with multimedia information. Such dictionaries have the potential to provide educational (Popescu et al., 2006), cross-langauge information retrieval (Hayashi et al., 2009) or assistive communication tools especially for children, language learners, speakers of different languages, and people with disabilities such as dyslexia (Mihalcea and Leong, 2008; Goldberg et al., 2009).

Additionally, a database of typical images connected to meanings has the potential to fill the gaps between images and meanings (semantic gap). There are many studies which aim to cross the semantic gap (Ide and Yanai, 2009; Smeulders et al., 2000; Barnard et al., 2003) from the point of view of image recognition. However the semantic classes of target images are limited (e.g. Caltech-101, 256[1]). Yansong and Lapata (2008) tried to construct image databases annotated with keywords from Web news images with their captions and articles, though the semantic coverage is unknown. In this paper, we aim to supply several suitable images for dictionary definitions. We propose a simple but effective method based on an Internet image search.

There have been several studies related to supplying images for a dictionary or thesaurus. Bond et al. (2009) applied images obtained from the Open Clip Art Library (**OCAL**) to Japanese **WordNet**.[2] They obtained candidate images by comparing the hierarchical structures of **OCAL** and **WordNet**, and then judged whether or not the image was suitable for the synset by hand. **OCAL** benefits from being in the public domain; however, it cannot cover a wide variety of meanings because of the limited number of available images.

Fujii and Ishikawa (2005) collected images and text from the Internet by querying lemma, and linked them to an open encyclopedia, **CYCLONE**.[3] They guessed the meaning of the images by disambiguating the surrounding text. This is a straightforward approach, but it is difficult to use it to collect images with minor meanings, because in most cases the Internet search querying lemma only provides images related to the most common meaning. For example, lemma アーチ *arch* may mean `''architecture''` or `''home run''` in Japanese, but a lemma search provided no image of the latter at least in the top 500.

There are some resources which link images to target synsets selected from **WordNet** (Fellbaum, 1998). For example, **PicNet** (Borman et al., 2005), **ImageNet** (Deng et al., 2009) and image ontology (Popescu et al., 2006, 2007; Zinger et al., 2006) collect candidate images from the Internet. **PicNet** and **ImageNet** ask Web users to judge their suitability, and Zinger et al. (2006); Popescu et al. (2007) automatically filtered out unsuitable images using visual characteristics. These approaches can

[1] http://www.vision.caltech.edu/Image_Datasets/Caltech101, 256/

[2] http://nlpwww.nict.go.jp/wn-ja/

[3] http://cyclone.cl.cs.titech.ac.jp/

Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 331–339, Beijing, August 2010

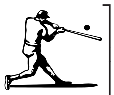
INDEX	アーチ *arch*	(POS: noun)	
SENSE 1	DEFINITION	上部$_1$ を 弓$_1$ の 形$_1$ に し た 建物$_1$ 。 また$_9$ 、 その$_3$ 建築$_1$ 様式$_2$ *Buildings with bow-shaped top. Or its architectural style.*	IMAGE 
	EXAMPLE	あの$_2$ 橋$_1$ は 2 つ の アーチ$_1$ で 出来$_4$ て いる 。 *That bridge has 2 arches.*	
	HYPERNYM	建物$_1$ *building*, 様式$_2$ *style*	
	SEM. CLASS	⟨`865:house (main building)`⟩ (⊂ ⟨`2:concrete`⟩), ⟨`2435:pattern, method`⟩ (⊂ ⟨`1000:abstract`⟩)	
SENSE 3	DEFINITION	野球$_1$ で 、 本塁打$_1$ 。 ホームラン$_1$ 。 *A home run in baseball.*	IMAGE 
	EXAMPLE	バッター$_1$ が ライト$_4$ スタンド$_2$ に 逆転$_3$ の アーチ$_3$ を 放っ$_4$ た *A batter blasted the ball over the right-field wall.*	
	HYPERNYM	本塁打$_1$ *honruida*	
	SYNONYM	ホームラン$_1$ *home run*, DOMAIN 野球$_1$ *baseball*	
	SEM. CLASS	⟨`1680:sport`⟩ (⊂ ⟨`1000:abstract`⟩)	

Figure 1: Simplified Entry for **Lexeed** & **Hinoki**:アーチ *arch*

collect a large number of highly accurate images. However, target synsets are limited at present, and the coverage of polysemous words is unknown. We present a comparison with **ImageNet** and image ontology (Popescu et al., 2006) in § 3.

In this paper, to cover a broad range of meanings, we use an Internet search. In advance, we expand the number of queries per meaning using information extracted from definition sentences. In § 3, we investigate the usability and effectiveness of several types of information targeting two different types of dictionaries, a Japanese Semantic Lexicon: **Lexeed** and a Web Dictionary: Japanese **Wikipedia**[4] (§ 2). We show that our method is simple but effective. We also analyze senses that are difficult to portray using images.

2 Resources

2.1 Japanese Semantic Lexicon: Lexeed

We use **Lexeed**, a Japanese Semantic Lexicon (Kasahara et al., 2004) as a target dictionary (see Figure 1). **Lexeed** includes the 29,000 most familiar words in Japanese, split into 48,000 senses. Each entry contains the word itself and its part of speech (POS) along with definition and example sentences and links to the Goi-Taikei (**GT**) Japanese Ontology (Ikehara et al., 1997). In addition, we extracted related words such as hypernyms, synonyms, and domains, from the definitions (called **Hinoki** Ontology). The images in Figure 1 are samples provided using our method.

[4]http://ja.wikipedia.org/

Table 1: Size of **Lexeed** and Japanese **Wikipedia** (disambiguation)

No.	**Lexeed**	**Wikipedia**	Shared Lemma
Entries	29,272	33,299	2,228
Senses	48,009	197,912[1]	19,703
Ave. Senses/Entry	1.6	5.9	8.8
Max. Senses/Entry	57	320	148
Monosemous	19,080	74	2
Ave. Words/Definition[2]	14.4	10.7	11.0

[1]From the all 215,883 lists, we extracted lists showing senses obtained by heuristics (see lines 2,3,4,6,7,9 and 10 for Figure 2).
[2]Analyzed by Mecab, http://mecab.sourceforge.net/

2.2 Web Dictionary :Japanese Wikipedia

We used **Wikipedia**'s disambiguation pages,[5] as a target dictionary (see Figure 2). A disambiguation page lists articles (eg. `‘‘European Union’’`, `‘‘Ehime University’’`) associated with the same lemma (eg. "EU"). Our goal is to provide images for each article listed. As shown in Figure 2, they include various writing styles.

2.3 Comparison of Lexeed and Wikipedia

Table 1 shows the sizes of **Lexeed** and **Wikipedia**'s disambiguation pages, and the shared entries. Shared entries are rare, and account for less than

[5]Version 20091011.

Original (in Japanese)

```
1  '''EU'''
2  * [[欧州連合]]
3  * [[Europa Universalis]]シリーズ - [[パラ
    ドックスインタラクティブ]]の[[歴史シミュレー
    ションゲーム]]
4  * [[愛媛大学]](Ehime University) - [[愛媛
    県]][[松山市]]にある日本の[[国立大学]]
5  '''Eu'''
6  * [[ユウロピウム]]の元素記号
7  * [[ユーフォニアム]] - 金管楽器
8 '''eu'''
9 * [[.eu]] - 欧州連合の[[国別ドメイン]]
10 * [[バスク語]]の[[ISO 639|ISO 639-1
    言語コード]]
```

Gloss

```
1 '''EU'''
2 * [[European Union]]
3 * [[Europa Universalis]] series - a [[histori-
  cal computer game]] by [[Paradox Interactive]]
4 * [[Ehime University]] - a [[National Univer-
  sity]] in [[Matsuyama]],[[Ehime Prefecture]]
5 '''Eu'''
6 * [[Europium]]'s chemical element symbol
7 * [[euphonium]] - a brass instrument
8 '''eu'''
9 * [[.eu]] - [[country-code top-level domain]]
   for the European Union
10 * [[ISO 639|ISO 639-1 language code]] of
   [[Basque]]
```

[[ ]] shows a link in **Wikipedia**. And we assign each line a number for easy citation.

Figure 2: Simplified Example of **Wikipedia**'s Disambiguation Page: "EU (disambiguation)"

10 % of the total [6][7]. As regards **Lexeed**, 16,685 entries (57 %) do not appear in any of **Wikipedia**'s lemmas, not only in disambiguation pages.[8]

As shown in Table 1, **Wikipedia** has many senses, but most of them are proper nouns. For example, in **Lexeed**, ヒマワリ *sunflower* is monosemous, but in **Wikipedia**, 67 senses are listed, including 65 proper nouns besides ``plant'' and ``sunflower oil''. On the other hand, in **Wikipedia**, アーチ *arch* has only one sense, ``architecture'' corresponding to **Lexeed**'s アーチ$_1$ *arch*, and has no disambiguation page.

As mentioned above, **Lexeed** and **Wikipedia** have very different types of entries and senses. This research aims to investigate the possibility of supplying appropriate images for such different senses, and a method for obtaining better images.

3 Experiment to Supply Images for Word Senses

In this paper, we propose a simple method for supplying appropriate images for each dictionary sense of a word. We collect candidate images from the Internet by using a querying image search. To obtain images even for minor senses, we expand the query by appending queries extracted from definitions for each sense.

In this paper, we investigated two main types of expansion, that is, the appending of mainly synonyms (SYN), and related words including hypernyms (LNK). For information retrieval, query expansion using synonyms has been adopted in several studies (Voorhees, 1994; Fang and Zhai, 2006; Unno et al., 2008). Our LNK is similar to methods used in Deng et al. (2009), but we note that their goal is not to give images to polysemous words (which is our intention). Popescu et al. (2006) also used synonyms (all terms in a synset) and hypernyms (immediate supertype in **WordNet**), but they did not investigate the effectiveness of each expansion and they forcus only on selected object synsets.

3.1 Experimental and Evaluation Method

We collected five candidate images for each sense from the Internet by querying an image search engine.[9] Then we manually evaluated the suitability of the image for explaining the target sense. The evaluator determined whether or not the image was appropriate (**T**), acceptable (**M**), or inappropriate (**F**). The evaluator also noted the reasons for **F**.

Figure 3 shows an example for たまねぎ *onion*. As shown in Figure 3, the evaluator determined **T**, **M** or **F** for each candidate image.

[6]Shared lemmas are そば *buckwheat noodle*, サイクル *cycle*, フクロウ *owl*, etc.

[7]Lemmas only in **Wikipedia** are イソップ *Aesop*, ビオ *Biot/Veoh*, 竜門の滝 *fall name*, etc.

[8]Lemmas only in **Lexeed** are 後払い *pay later*, ユーモラス *humorous*, 抜擢 *selection*, etc.

[9]We used Google AJAX images API, http://code.google.com/intl/ja/apis/ajaxsearch/

(1)  (2)  (3)  (4)  (5) 

T (Appropriate) **F** (Inappropriate) **M** (Acceptable) **T** (Appropriate) **T** (Appropriate)

Figure 3: Examples of Candidate Images and Evaluations for たまねぎ *onion*

Table 2: Data for **Hinoki** Ontology

Type	No.	%	Example	
			Lemma	Related Word
Hypernym	47,054	69.1	アーチ$_1$ *arch*	様式
Synonym	14,068	20.6	アーチ$_3$ *arch*	ホームラン *homer*
Domain	1,868	2.7	アーチ$_3$ *arch*	野球 *baseball*
Hyponym	757	1.1	売り買い$_1$ *buy and sell*	売る *sell*
Meronym	686	1.0	赤身$_1$ *lean*	魚肉 *fish meat*
Abbreviation	383	0.6	亜$_2$ *A(sia)*	アジア *Asia*
Other name	216	0.3	差し込み$_2$ *shave*	コンセント *plug outlet*
Other	3102	4.6	包み焼き$_1$ *papillote*	魚 *fish*
Total	68,134	100		

For an image that is related but that does not explain the sense, the evaluation is **F**. For example, for たまねぎ *onion*, the images of onion dishes such as (2) in Figure 3 are **F**. On the other hand, the images that show onions themselves such as (1), (4) and (5) in Figure 3 are **T**. With (3) in Figure 3, the image may show the onion itself or a field of onions, therefore the evaluation is **M**.

One point of judgment, specifically between **T** and **M**, is whether the image is typical or not. With たまねぎ *onion*, most typical images are similar to (1), (4) and (5). The image (3) may not be typical but is helpful for understanding, and (2) may lead to a misunderstanding if this is the only image shown to the dictionary user. This is why (3) is judged to be **M** and (2) is judged to be **F**.

We evaluated 200 target senses for **Lexeed**, and 100 for **Wikipedia**.[10]

3.2 Experiment: Lexeed

In this paper, we expand queries using the **Hinoki** Ontology (Bond et al., 2004), which includes related words extracted from the definition sentences. Table 2 shows the data for the **Hinoki** Ontology.

For SYN, we expand queries using synonyms, abbreviations, other names in Table 2, and variant spellings found in the dictionary. On the other hand, for LNK, we use all the remaining relations, namely hypernyms, domains, etc. Additionally, we use only normal spellings with no expansion, when the target words are monosemous (MONO). One exception should be noted. When the normal spelling employs hiragana (Japanese syllabary characters), we expand it using a variant spelling. For example, とんぼ *dragonfly* is expanded by the variant spelling 蜻蛉 *dragonfly*.

To investigate the trends and difficulties based on various conditions, we split the **Lexeed** senses into four types, namely, concrete and monosemous (MC), or polysemous (PC), not concrete and monosemous (MA), or polysemous (PA). We selected 50 target senses for evaluation randomly for each type. The target senses were randomly selected without distinguishing them in terms of their POS.

Note that we regard the sense as being something concrete that is linked to **GT**'s semantic classes subsumed by ⟨`2:concrete`⟩, such as たまねぎ *onion* (⊂ ⟨`677:crop/harvest/farm products`⟩ ⊂ ⟨`2:concrete`⟩).

3.3 Results and Discussion: Lexeed

Table 3 shows the ratio of **T** (appropriate), **M** (acceptable) and **F** (inappropriate) images for the target sense. We calculated the ratio using all five candidate images, for example, in Figure 3, the

[10] We performed an image search in September 2009 for **Lexeed**, and in December 2009 for **Wikipedia**.

ratio of appropriate images is 60 % (three of five).

In Table 3, the baseline shows a case where the query only involves the lemma (normal spelling). As shown in Table 3, SYN has higher precision than LNK. This means that SYN can focus on the appropriate sense. With polysemous words (PC, PA), expansion works more effectively, and helps to supply appropriate images for each sense. However, with MC, both LNK and SYN have less precision. This is because the target senses of MC are majorities, so expansion is adversely affected. Although MONO alone has good precision, because hiragana is often used as readings and has high ambiguity, appending the variant spelling helps us to focus on the appropriate sense.

Here, we focus on LNK of PC, and then analyze the reasons for **F** (Table 5). In Table 5, in 24.3% of cases it is "difficult to portray the sense using images" (The numbers of senses for which it is "difficult to portray the sense using images" are, 3 of MC, 9 of PC, 10 of MA, and 16 of PA. We investigate such senses in more detail in § 3.4.).

For such senses, no method can provide suitable images, as might be expected. Therefore, we exclude targets where it is "difficult to portray the sense using images", then we recalculated the ratio of appropriate images. Table 4 shows the capability of our proposed method for senses that can be explored using images. This leads to 66.3 % precision (15.3% improvement) even for most difficult target type, PA.

Again, when we look at Table 5, reasons 2-5 (33.3 %) will be improved. In particular, "hypernym leads to ambiguity" makes up more than 10%. Hypernyms sometimes work well, but sometimes they lead to other words included in the hypernyms. For example, appending the hypernym 食品 *foods* to 煮干し *boiled-dried fish* leads to images of "foods made with boiled-dried fish". This is why SYN obtained better results than LNK. Then, with "expanded by minor sense" and when the original sense is dominant majority, expansion reduced the precision. Therefore, we should expand using only words with major senses.

3.4 Discussion: Senses can/cannot be shown by images

As described above, the target senses are randomly selected without being distinguished by their POS, because we also want to investigate the features of senses that can be shown by images. Table 6 shows the ratio of senses judged as "difficult to portray the sense using images" (labeled as "Not Shown") for each POS. As regards POS, the majority of selected senses are nouns, followed by verbal nouns and verbs. We expected that the majority of nouns and verbal nouns whould be "Shown", but did not expect that a majority of verb is also "Shown". Other POSs are too rare to judge, although they tend to fall in the "Not Shown" category.

Furthermore, in Table 7, for nouns and verbal nouns, we show the ratio of senses for each type ("Concrete" or "not Concrete") judged in terms of "difficult to portray the sense using images". We classified the senses into "Concrete" or "not Concrete" based on **GT**'s semantic classes, as described in § 3.2.

Table 6: Ratio of Senses judged as "difficult to portray the sense using images" for each POS

POS	Shown		Not Shown		Total
	No.	%	No.	%	No.
Noun	132	85.2	23	14.8	155
Verbal Noun	15	78.9	4	21.1	19
Verb	9	81.8	2	18.2	11
Affix	4	57.1	3	42.9	7
Pronoun	0	0	2	100	2
Adjective	1	50	1	50	2
Adverb	0	0	2	100	2
Interjection	1	100	0	0	1
Conjunction	0	0	1	100	1
Total	162	81	38	19	200

Table 7: Ratio of Concrete/Not Concrete Senses judged as "difficult to portray the sense using images": for Nouns and Verbal Nouns

Type	Shown		Not Shown		Total
	No.	%	No.	%	No.
Concrete	114	90.5	12	9.5	126
Not Concrete	33	68.8	15	31.3	48
Total	147	84.5	27	15.5	174

Table 3: Ratio of Appropriate Images for Sense (Precision): **Lexeed**

Target Type		Expanding Method	**F** (Inappropriate) No.	%	**T** (Appropriate) No.	%	**M** (Acceptable) No.	%	**T+M** No.	%	Total
Con-	Mono-	SYN	18	24.0	36	48.0	21	28.0	57	76.0	75
	semous	LNK	82	**33.5**	112	45.7	51	20.8	163	66.5	245
	(MC)	MONO	42	16.8	181	**72.4**	27	10.8	208	**83.2**	250
		baseline	46	18.4	171	68.4	33	13.2	204	81.6	250
crete	Poly-	SYN	94	38.7	88	36.2	61	25.1	149	**61.3**	243
	semous	LNK	111	**44.4**	92	36.8	47	18.8	139	55.6	250
	(PC)	baseline	180	**72.0**	53	21.2	17	6.8	70	28.0	250
not	Mono-	SYN	32	42.7	21	28.0	22	29.3	43	**57.3**	75
	semous	LNK	138	**57.5**	54	22.5	48	20.0	102	42.5	240
Con-	(MA)	MONO	98	40.0	98	40.0	49	20.0	147	**60.0**	245
		baseline	112	44.8	86	34.4	52	20.8	138	55.2	250
crete	Poly-	SYN	122	49.0	64	25.7	63	25.3	127	**51.0**	249
	semous	LNK	150	**60.2**	52	20.9	47	18.9	99	39.8	249
	(PA)	baseline	201	**80.7**	36	14.5	12	4.8	48	19.3	249

Table 4: Ratio of Appropriate Images for Sense (Precision), excluding senses that are difficult to portray using images: **Lexeed**

Target Type		Expanding Method	**F** (Inappropriate) No.	%	**T** (Appropriate) No.	%	**M** (Acceptable) No.	%	**T+M** No.	%	Total
Con-	Mono-	SYN	15	**21.4**	36	51.4	19	27.1	55	**78.6**	70
	semous	LNK	71	**30.9**	112	48.7	47	20.4	159	69.1	230
	(MC)	MONO	29	12.3	180	76.6	26	11.1	206	**87.7**	235
		baseline	35	14.9	170	72.3	30	12.8	200	85.1	235
crete	Poly-	SYN	61	30.8	85	42.9	52	26.3	137	**69.2**	198
	semous	LNK	84	40.0	89	42.4	37	17.6	126	60.0	210
	(PC)	baseline	139	67.8	53	25.9	13	6.3	66	32.2	205
not	Mono-	SYN	17	34.0	20	40.0	13	26.0	33	**66.0**	50
	semous	LNK	101	**51.8**	54	27.7	40	20.5	94	48.2	195
Con-	(MA)	MONO	65	33.3	94	48.2	36	18.5	130	**66.7**	195
		baseline	72	36	85	42.5	43	21.5	128	64.0	809
crete	Poly-	SYN	57	33.7	63	37.3	49	29	112	**66.3**	169
	semous	LNK	81	47.9	52	30.8	36	21.3	88	52.1	169
	(PA)	baseline	122	72.2	36	21.3	11	6.5	47	27.8	169

Table 5: Reasons for **F**: PC, LNK:**Lexeed**

No.	Reason	No.	%	Example
1	difficult to portray the sense using images	27	24.3	これ *me* `‘‘humble expressions used for oneself’’`
2	hypernym leads to ambiguity	12	10.8	煮干し *boiled-dried fish* (⊂ 食品 *foods*)
3	expanded by minor sense	11	9.9	リンク *link* (⊂ リンクス *links*, usually means *lynx*)
4	no expansion is better	8	7.2	カメラマン *cameraman* (⊂ 部員 *staff*)
5	original sense is TOO minor	6	5.4	海 *lake* (⊂ 湖 *lake*), 海 usually means *sea*
6	Other	47	42.3	
	Total	111	100	

As shown in Table 7, 90.5 % of "Concrete" nouns are judged as "Shown", and only 9.5 % of senses are judged as "Not Shown" [11]. However 68.8 % of "not Concrete" nouns are also judged as "Shown".

Therefore, both POS and type ("Concrete" or "not Concrete") are helpful, but not perfect features as regards knowing the sense is "difficult to portray the sense using images". In future work we will undertake further analysis to determine the critical features.

3.5 Experiment: **Wikipedia**

For LNK we use the **Wikipedia** hyperlinks (shown as [[ ]] in Fig 2). 95.5 % of all senses include [[ ]], 85.4 % linked to an actual page, and [[ ]] appeared 0.95 times per sense. Note that we do not use time expression links such as [[2010]] and [[1990s]].

With SYN, we use synonyms extracted with heuristics. Table 8 shows the main rules that we used to extract synonyms. We extracted synonyms for 98.0 % of 197,912 senses.

Then we randomly selected 50 target senses for evaluation from lemmas shared/unshared by **Lexeed**.

3.6 Results and Discussion: **Wikipedia**

We do not show the baseline in Table 9, but it is always below 10%. For all target senses, expansion provides more suitable images. Because there are so many senses in **Wikipedia**, no target sense is in the majority. As shown in Table 9, there are few differences between SYN and LNK, because most of the synonyms used for SYN are also links. However, SYN has slightly superior precision as regards **T** (Appropriate), which means the process of extracting synonyms helped to reject links that were poorly with the target senses.

Also in **Lexeed**, expansion using synonyms (SYN) had higher precision than hypernyms (LNK). Because we do not know the total number of suitable images for the target senses on the Internet, we cannot estimate the recall with this evaluation method. However, we speculate that hypernyms provide higher recall. Deng et al. (2009) undertook expansion using hypernyms and this may be an appropriate way to obtain many more images for each sense. However, because our aim is employ several suitable images for each sense, high precision is preferable to high recall.

Now, we focus on LNK shared by **Lexeed**, and then we analyze the reasons for **F** (Table 10). In contrast to **Lexeed**, no sense is classified as "difficult to portray the sense using images". However, there are many senses where it is difficult to decide what kind of images "explain the target sense". For example, in Table 10, with "maybe **T** (Appropriate)", the target sense was a personal name and the image was his/her representative work. In this paper, for personal names, only the images of the person are judged to be **T**, despite the fact that supplying images of representative work for novelists or artists may be suitable.

In this study, we obtained five images per sense, but only one image was sufficient for some senses, for example, an image of an album cover for the name of an album. In contrast, several different types of images are needed for some senses. For example, for the name of a city, images of maps, landscapes, city offices, symbols of the city, etc. are all suitable. Therefore, it may be better to estimate a rough class first, such as the name of an album, artist and place, and then obtain preassigned types of images.

4 Conclusions

The goal of this work was to supply several suitable images for dictionary definitions. The target dictionaries were **Lexeed** and **Wikipedia**, which have very different characteristics. To cover a wide range of senses, we collected candidate images from the Internet by querying an image search engine. Then, to obtain suitable and different images for each sense, we expanded the queries by appending related words extracted from the definition sentences. In this paper, we tried two types of expansion, one mainly using synonyms (SYN), and one mainly using hypernyms or related links (LNK).

The results show that SYN provided better precision than LNK, especially for **Lexeed**. Also, query expansion provided a substantial improvement for

[11] For example, 学会 *conference* (⊂ ⟨373:organization, etc.⟩ ⊂ ⟨2:concrete⟩), 親代わり *parental surrogate* (⊂ ⟨342:agent/representative⟩ ⊂ ⟨2:concrete⟩), and so on.

Table 8: Rules for Extracting Synonyms for SYN: **Wikipedia**

Rule	Example: Lemma	Definition sentences
head parts separated by hyphen (- or –)	EU	[[euphonium]] - a brass instrument (line 7 in Figure 2)
whole definitions appear as a chunk	EU	[[European Union]] (line 2 in Figure 2)
parts indicated by		
arrow (→)	イヌ *dog*	One of [[Oriental Zodiac]]→[[戌 *dog*]]
quotation key words, 参照 *See* etc.	イヌ *dog*	[[Chinese character]]'s [[radical parts]], See [[犬部 *inu-bu*]]
parts in parentheses or " " including		
whole lemma	Einstein	"Albert Einstein"
alphameric characters, for katakana lemma	サンバ	"samba"
characters of alpha-numeral lemma	CS	コンピュータ科学 (computer science)

underlined parts show the extracted synonyms.

Table 9: Ratio of Appropriate Images for Sense (Precision): **Wikipedia**

Target Type	Expanding Method	**F** (Inappropriate) No.	%	**T** (Appropriate) No.	%	**M** (Acceptable) No.	%	**T+M** No.	%	Total
Shared by **Lexeed**	SYN	98	40.8	119	**49.6**	23	9.6	142	**59.2**	240
	LNK	92	41.8	107	48.6	21	9.5	128	58.2	220
NOT shared by **Lexeed**	SYN	100	41.2	103	**42.4**	40	16.5	143	58.8	243
	LNK	96	41.0	93	39.7	45	19.2	138	**59.0**	234

Table 10: Reasons for **F**: Shared by **Lexeed**, LNK: **Wikipedia**

No.	Reason	No.	%	Example: Lemma	Links
7	lack of queries (available words in def.)	14	15.2	ふえ *fue (reading)*	フエ *Hue, city name in Vietnam*
8	inappropriate queries (available words in def.)	10	10.9	レギュラー *regular*	出場選手登録 *active roster*
2	hypernym lead to ambiguity	5	5.4	キャッシュ *cache*	ジオキャッシング *geocaching*
9	maybe **T** (Appropriate)	5	5.4	モンキー *monkey*	モンキー・パンチ *Monkey Punch*
6	Other	58	63		
	Total	92	100		

polysemous words. Our proposed method is simple but effective for our purpose, that is supplying suitable and different images for each sense.

In future work we intend to analyze senses that are difficult/easy to portray using images in more detail, using not only semantic charactaristics but also visual features(Csurka et al., 2004). We also intend to improve the expansion method. One way to achieve this is to filter out expansions with minor senses. As for **Wikipedia**, we should approximate the class first, such as the name of an album, artist and place, then obtain preassigned types of images.

References

Kobus Barnard, Pinar Duygulu, Nando de Freitas, David Forsyth, David Blei, and Michael I. Jordan. 2003. Matching Words and Pictures. *Journal of Machine Learning Research*, Vol. 3, pp. 1107–1135.

Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. 2009. Enhancing the Japanese WordNet. In *The 7th Workshop on Asian Language Resources, in conjunction with ACL-IJCNLP-2009*, pp. 1–8.

Francis Bond, Eric Nichols, Sanae Fujita, and Takaaki Tanaka. 2004. Acquiring an Ontology for a Fundamental Vocabulary. In *Proceedings of the 20th International Conference on Computational Linguistics: COLING-2004*, pp. 1319–1325.

Andy Borman, Rada Mihalcea, and Paul Tarau. 2005. PicNet: Pictorial Representations for Illustrated Semantic Networks. In *Proceedings of the AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors*.

Gabriela Csurka, Cedric Bray, Chris Dance, and Lixin Fan. 2004. Visual categorization with bags of keypoints. In *ECCV International Workshop on Statistical Learning in Computer Vision*, pp. 59–74.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *IEEE Computer Vision and Pattern Recognition (CVPR)*.

Hui Fang and ChengXiang Zhai. 2006. Semantic term matching in axiomatic approaches to information retrieval. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pp. 115–122. ACM.

Christine Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Atsushi Fujii and Tetsuya Ishikawa. 2005. Image Retrieval and Disambiguation for Encyclopedic Web Search. In *Proceedings of the International Joint Conference on Artificial Intelligence: IJCAI-2005*, pp. 1598–1599.

Andrew B. Goldberg, Jake Rosin, Xiaojin Zhu, and Charles R. Dyer. 2009. Toward Text-to-Picture Synthesis. In *NIPS 2009 Mini-Symposia on Assistive Machine Learning for People with Disabilities*.

Yoshihiko Hayashi, Savas Bora, and Masaaki Nagata. 2009. Utilizing Images for Assisting Cross-language Information Retrieval on the Web. In *International Workshop on Web Information Retrieval Support Systems*, pp. 100–103.

Ichiro Ide and Keiji Yanai. 2009. Crossing the Semantic Gap : Towards the Understanding of Image and Video Contents. *Journal of Japanese Society for Artificial Intelligence*, Vol. 24, No. 5, pp. 691–699. (in Japanese).

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. 1997. *Goi-Taikei — A Japanese Lexicon*. Iwanami Shoten, Tokyo. 5 volumes/CD-ROM.

Kaname Kasahara, Hiroshi Sato, Francis Bond, Takaaki Tanaka, Sanae Fujita, Tomoko Kanasugi, and Shigeaki Amano. 2004. Construction of a Japanese Semantic Lexicon: Lexeed. In *IEICE Technical Report: 2004-NLC-159*, pp. 75–82. (in Japanese).

Rada Mihalcea and Chee Wee Leong. 2008. Toward communicating simple sentences using pictorial representations. *Machine Translation*, Vol. 22, No. 3, pp. 153–173.

Adrian Popescu, Christophe Millet, and Pierre-Alain Moëllic. 2007. Ontology Driven Content Based Image Retrieval. In *Proceedings of the ACM International Conference on Image and Video Retrieval*.

Adrian Popescu, Christophe Millet, Pierre-Alain Moëllic, Patrick Hède, and Gregory Grefenstette. 2006. Automatic Construction of a Grounded Multimedia Ontology of Objects to Illustrate Concepts in a Learning Process. In *NETTIES 2006 Conference: Advanced Educational Technologies for a Future e-Europe*.

Arnold W.M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. 2000. Content-based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1349–1380.

Yuya Unno, Yusuke Miyao, and Jyunichi Tujii. 2008. Information Retrieval using Automatically Extracted paraphrases. In *Proceedings of the 14th Annual Meeting of The Association for Natural Language Processing: NLP-2008*, pp. 123–126. (in Japanese).

Ellen M. Voorhees. 1994. Query Expansion using Lexical-Semantic Relations. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pp. 61–69.

Feng Yansong and Mirella Lapata. 2008. Automatic image annotation using auxiliary text information. In *Proceedings of ACL-08: HLT*, pp. 272–280. Association for Computational Linguistics.

Svitlana Zinger, Christophe Millet, Benoit Mathieu, Gregory Grefenstette, Patrick Hède, and Pierre-Alain Moëllic. 2006. Clustering and semantically filtering web images to create a large-scale image ontology. In *SPIE 18th Annual Symposium Electronic Imaging, Internet Imaging VII*, Vol. 6061, pp. 89–97.