# PicNet: Augmenting Semantic Resources with Pictorial Representations

**Andy Borman, Rada Mihalcea, Paul Tarau**
Department of Computer Science and Engineering
University of North Texas
{rada,dab0068,tarau}@cs.unt.edu

**Abstract**

In this paper we introduce PICNET – a Web-based system for augmenting semantic resources with illustrative images using volunteer contributions over the Web. PICNET seeks to build rich knowledge-bases that encode word/image associations, to the end of combining the advantages and power of both visual and linguistic representations as means of defining world concepts. In this paper, we address some of the issues encountered in identifying prototypical illustrations for various concepts, as well as issues related to the construction of such pictorial knowledge-bases with the help of Web users.

## Introduction

The capacity to communicate our day-by-day experiences stands mainly in the capacity to describe in a meaningful way the entities we deal with in our everyday life, as well as the interactions between such entities. While there are a variety of forms of communication that people use to transfer meaningful information from one another (e.g. gaze or eye contact, touch, paralanguage attributes such as sighs or voice tone, etc.), verbal and visual expressions are perhaps the two most often encountered means of communication, often times used together to convey information.

The conversion of such forms of communication into formats that can be stored electronically – and thus made suitable for further automatic processing, results into: (1) written or spoken text – as a representation of verbal communication, and (2) images or videos – as means of representing visual information. While these two forms of communications are usually inseparable in the way they are used by humans, current research in language and image processing has not yet reached the point where verbal and visual computer-based representations interact in any meaningful ways.

In the PICNET project, we are concerned with the dual verbal/visual representations of words and word phrases – seen as cognitive units representing the smallest units of communication. Starting with a machine readable dictionary that defines the words in the common vocabulary and their possible meanings, we seek to add visual representations to the dictionary entries, to the end of building a knowledge-base that combines both verbal and visual representations of these basic cognitive units.

*"A picture is worth a thousand words."* The main idea behind PICNET is to build illustrated semantic networks that encode *word/image* associations, combining together the advantages and power of both linguistic and visual representations as means of defining world concepts. This opens the use of lexical knowledge-bases like WordNet to a larger user base comprising children, non-English speakers, people with disabilities or limited reading skills, while extending the power of semantic networks to richer knowledge-bases that include word/image and image/image associations.

Since we are interested in the association of words with images, all language speakers are, by definition, experts in this field. We are therefore trying to build this resource using a Web-based system, which allows Web users to contribute their knowledge to encoding or validating word/image associations.

In this paper, we present the paradigm behind PICNET, the implementation and the interface of this Web-based system, and the role of Web users in contributing and validating pictorial representations for concepts defined in a semantic network. We also introduce a new scoring scheme that jointly evaluates contributions made by users in different activities, and provides a unified ranking indicating the overall quality of the contributions.

## Motivation

Children learn language by being exposed to visual representations of the new entities or actions they interact with. Similarly, communication with persons having a different linguistic background is often done through signs or drawings that offer alternative renderings of the information we want to convey. Through its word/image associations, PICNET has the capability to become a natural catalyst for bringing the knowledge encapsulated in semantic networks like WordNet to a larger user community. We hope that language learning, for children or for people with language disorders, will genuinely benefit from the connections between meanings of words and their visual representations, which are made explicit within the PICNET knowledge-base.

PICNET can be also seen as an associative facilitator for people learning a second language, as well as an *international* language-independent knowledge-base. While understanding the word *house* requires knowledge of English, understanding a picture that represents a *house* is transparent to languages – and such a representation can be understood by speakers of any language, regardless of their origins or literacy[1]. As a derived benefit, by exploiting the language-independent aspect of visual representations, PICNET could provide the means for building multi-lingual dictionaries and semantic knowledge-bases, without requiring bilingual speakers. For instance, by showing an image associated with a given concept and asking users to provide the corresponding words in the various languages, can result into a self-managed multi-lingual extension mechanism for knowledge-bases like WordNet.

Moreover, PICNET has also the potential to bridge the gap between research work in image processing and language processing. For instance, the explicit word/image associations available in PICNET, as well as the hierarchical relations among pictures automatically derived from the corresponding semantic network (WordNet) can help improve the quality of systems for image retrieval and/or classification. Similarly, image content analysis can help language processing tasks. For example, (Barnard, Johnson, & Forsyth 2003) showed how text and image information can be jointly used to improve the performance of a word sense disambiguation task. Other tasks, such as information extraction, information retrieval, named entity recognition, etc., could be also improved using dual visual/linguistic representations.

## Representing Word/Image Associations

The first iteration of PICNET focuses on concrete nouns and noun-phrases, which can be easily represented using an image. Nouns such as *African violet* or *dachshund* are examples of English lexicalizations for concepts that correspond to concrete entities in the real world. There are however other types of entities (e.g. general or abstract nouns), as well as interactions or various entity properties that might require different visual representations. For instance, the image associated with a general noun such as *flower* should integrate a range of instances of type *flower*, to suggest an entire category rather than one specific instance. Similarly, abstract nouns can only be defined based on common sense associations – e.g. *philosophy* could be represented with faces of well-known philosophers, *politics* with faces of well-known politicians, etc. In this section, we discuss a few possible mechanisms for providing visual representations for words other than concrete nouns, although extensive testing is required to demonstrate their applicability in practice.

**Generalization as Instance Morphing.**
An intuitive way to represent general concepts is through a morphing sequence of some of their most typical instances. WordNet hyponym links can be used to provide such instances, together with frequency information. For instance, the abstract concept *car* could be represented as a morphing sequence cycling over a *sedan*, *van*, *SUV* and a *sports car*. A *feline* can be represented as a morphing of typical instances like *cats*, *tigers*, *pumas* etc. Such sequences can be obtained automatically and can be displayed as animated GIF files or other portable animation formats.

**Abstraction via Common-Sense Associations.**
Abstract nouns, such as e.g. *philosophy*, *friendship*, do not usually correspond to physical entities, and thus they do not admit direct visual representations. Instead, other concrete nouns that often co-occur with these abstractions can be used to render in a graphical way the information conveyed by these abstract concepts. For instance, an image of *Socrates*, or a morphing sequence cycling over images of *Socrates*, *Plato*, and famous groups or schools of philosophers – can serve as a representation for the abstract concept of *philosophy*. Similarly, *friendship* can be represented as a sequence of images illustrating people (or other beings) that are close together, *love* could be represented as a set of images showing concrete entities that are commonly used to illustrate this concept, such as *heart*, *wedding*, *Eros*.

**Representing Properties.**
While adjectives denoting perceptual properties like colors can be represented directly as samples, more complex properties can be described as collections of concrete entities that posses them.

**Representing Verbs.**
Verbs representing relations/predicates can be represented as graphs connecting their arguments with semantic roles as links. For instance, verbs expressing creation or change can be represented as short animated stories or video clips describing their typical semantic frames, using structures and relations as obtained through semantic parsing.

## Resources

The PICNET project aims to build large knowledge-bases of word/image associations with the help of Web users. One of the most important objectives targeted by the PICNET system design is to facilitate the task of these non-expert contributors as much as possible. That is, rather than asking the user to look for external resources for images or definitions, we try to provide several such resources directly on the system Web site. With such resources linked directly from the PICNET page, the task of the users is greatly simplified – they select the right information from a pool of readily available information and usually do not have to seek additional resources.

### WordNet

WordNet 2.0 is the primary information source that we use in PICNET for the construction of an illustrated semantic network. WordNet is a Machine Readable Dictionary developed at Princeton University by a group led by George Miller (Miller 1995), (Fellbaum 1998).

[1] Note that PICNET does not limit the number of images that can be associated with a given concept. It is not expected that one image can always fully represent a particular concept, but a set of images taken together may represent the collective notion of diverse contributors, and provide an expanded understanding to a user of the system.

WordNet covers the vast majority of nouns, verbs, adjectives and adverbs from the English language. The words in WordNet are organized in synonym sets, called *synsets*. Each synset represents a concept. WordNet 2.0 includes a large network of 152,050 words, organized in 115,424 synonym sets, called *synsets*. Table shows the number of nouns, verbs, adjectives and adverbs defined in WordNet 2.0, and the number of synsets for each of these parts of speech.

| Part of speech | Words | Synsets |
|---|---|---|
| Noun | 114,648 | 79,689 |
| Verb | 11,306 | 13,508 |
| Adjective | 21,436 | 18,563 |
| Adverb | 4,660 | 3,664 |
| TOTAL | 152,050 | 115,424 |

Table 1: Words and synsets in WordNet 2.0

While the current implementation of PICNET focuses on nouns only, in future work we plan to explore possible pictorial representations for other word classes, such as verbs, adjectives, or adverbs.

WordNet also includes a large number of semantic relations defined across concepts, including hypernymy/hyponymy (ISA), meronymy/holonymy (HASA), antonymy, etc. Note that semantic relations are defined among *concepts*, and not among *words*, and therefore we believe that such semantic links can be used to navigate or exploit in other meaningful ways the *network* of pictures built in PICNET.

### Image Search Engines

PICNET is seeded with images culled from automated PicSearch (`http://www.picsearch.com`) and AltaVista (`http://www.altavista.com/image/`) image searches. To date, 72,968 images have been automatically collected. Currently, the collected images are validated by Web volunteers, and we are also looking into automatic and semi-automatic validation mechanisms, involving image ranking algorithms, relying on iterative graphical models as proposed in previous work (Mihalcea & Tarau 2004).

This automated process is about 61% effective at finding a reasonable mapping[2]. The results of the automatic process were particularly good when searching for concrete nouns or specific entities with precise definitions, such as a particular plant genus. However, in general, the value of the automated search results is mixed. It seems that the search engines' procedure relies more on the image filename rather than the textual context of the image. Also, due to the sheer quantity of synsets given and the search and processing time required, no particular attempt was made to differentiate between different senses of a word when performing the automated seeding.

[2]The effectiveness was estimated through random sampling of user validated images. About 15% of the seedings were images that failed to copy (blank images). Another 5% were not useful (corporate logos, or images that were too small). Of the remainder, users indicated that 44% of the images were very good matches, while 17% were a near match, sharing some attributes in common with the synset. The final 19% were rejected by users as having nothing in common with the assigned synset.

## Activities in PICNET

The value of Web data collection systems lies in the nearly limitless availability of users that possess the knowledge the system aims to acquire. The PICNET Web interface aims to maximize this benefit by providing engaging challenges, competition, and constant feedback to keep the user engaged and motivated. An administrative facility allows the superuser to disallow certain uploaded images before they are incorporated into the system, and to undo the contributions of a particular user, if necessary.

### Searching

A user may search the PICNET dictionary for a word/image association, using PICNET for its intended purpose. The synsets are scanned for the keyword, and a list of results is displayed. Some synsets may not have an image associated with them yet, and the user may choose to exclude these synsets from the results. From the list of results, the user may, if dissatisfied with the results, elect to upload a new image for their selected synset, or simply comment on the quality of the image association by voting through the PICNET validation method (see details below).

The search is not strict, meaning that all synsets including the keyword indicated by the user will be included in the results. For example, a search for *tiger* will also return *tiger lily* and *tiger shark*.

A synset may be associated with several images, in which case the top few pairs are shown, where the quality of a pair is determined by the total value of user votes gleaned from the various PICNET processes. In many cases, a set of images can better encompass the meaning of the synset than a single result.

### Donating Images

A user may upload images for usage within the system. Images contributed in this manner are not immediately associated with a synset, but will gain their word associations in other PICNET phases.

### Free Association

In this task, the user is shown a random image from the dictionary. The image may or may not already have a synset association, which is not apparent to the user in this phase. The user is asked to enter a word related to the image. Figure 1 shows a snapshot of the the PICNET free association screen. The dictionary is searched and the user is asked to select the precise word sense intended. A new synset/image association is created (and given an automatic `+3` point vote).

It is possible that different users come up with entirely different suggestions for the same image. This is expected, as it may represent a relationship between the synsets that


Figure 1: Free-association screen in PICNET

PICNET predicts (such as hyponym-hypernym), or a relationship not catalogued in WordNet.

As another consideration, a user may be presented with an image which is more precise than her knowledge. For instance, when presented with a good image for *mule deer*, the user may know only that it is in the *deer* family. Incorporating this information is useful, as it verifies the hypernym-hyponym relationship. A future user exploring the PICNET mapping for *deer*, seeing the *mule deer* image along with images of other related sub-species, will better understand that *deer* is a grouping of a more specific subclass.

Through the process of image validation and dictionary commentary, it is expected that the precision of an image will come to match the precision of the synset definition.

## Validating Images

Synset/image associations are created by user uploads, user free association, PICNET guesses, or the initial automated PICNET seeding. The quality of these associations can be determined by the commentary of diverse users. PICNET uses a voting process to record this commentary.

In this task, users are shown a synset/image association drawn from the PICNET dictionary. The synset/image pair is drawn at random from the pool of associations having the fewest votes. However, associations which the user has previously voted on are excluded as are pairs which have received sufficient negative feedback. The user is shown the synset with its gloss (or definition) and the associated image. The user may then vote on the appropriateness of the image using the following options: this image (a) is NOT related to this concept; (b) is loosely related to this concept; (c) is related to many attributes of this concept; (d) is well related to this concept. The vote is recorded, the pair is scored, and the user is shown a new pair. Figure 2 shows a snapshot of the PICNET validation screen.

## Competitive Free Association

To help motivate users to participate, a game process was implemented to allow users to compete with each other. The game concept is loosely based on a combination of Out of the Box's Apples To Apples® game where players match their noun cards with their opponent's adjective card, the old Balderdash®-style dictionary game, and the PICNET's Free Association.

The rules of the game are simple. The game begins once a minimum of five players join and a majority votes to start. To start the round, each player is shown an image from the PICNET database and asked to provide an anonymous word association. Identical entries from multiple players are coalesced. After all players have entered a suggestion, the players vote for the best selection which is not their own. The player who entered the word receiving the most votes wins the round. If multiple players entered the winning word, the points for the round are split. The word/image associations are added to the PICNET database, as in the Free Association task, with the score value equal to the number of votes received.

The game is played in steps over several days, so users can login at their convenience and take their action. To keep the game advancing when a player has failed to take a turn, PICNET will spoof the human player, making an entry for him, using scoring information, educated guessing, and the WordNet relations in the PICNET database.

Response to the game was positive, and the results demonstrate some positives and pitfalls with the mechanism. In many cases, the players did not have the necessary knowledge to provide a synset answer matching the specificity of the image. For example, the players presented with a diagram of the *humerus* (a bone of the arm), knew only that it was a bone, or mistook it for a *leg bone*. To alleviate this problem, synsets drawn from PICNET's existing mappings are now included among the voting options. Since some of these mappings were created through the seeding process or through image assignment by an expert user, a mapping with appropriate specificity will be presented with the hope that users will elect the more precise mapping for this image.

## Other User Activities (work in progress)

WordNet provides glosses associated to each synset (definitions and examples of use). By providing a "semantic cloud" of image and textual links obtained after parsing such glosses, one can extend navigation capabilities beyond the

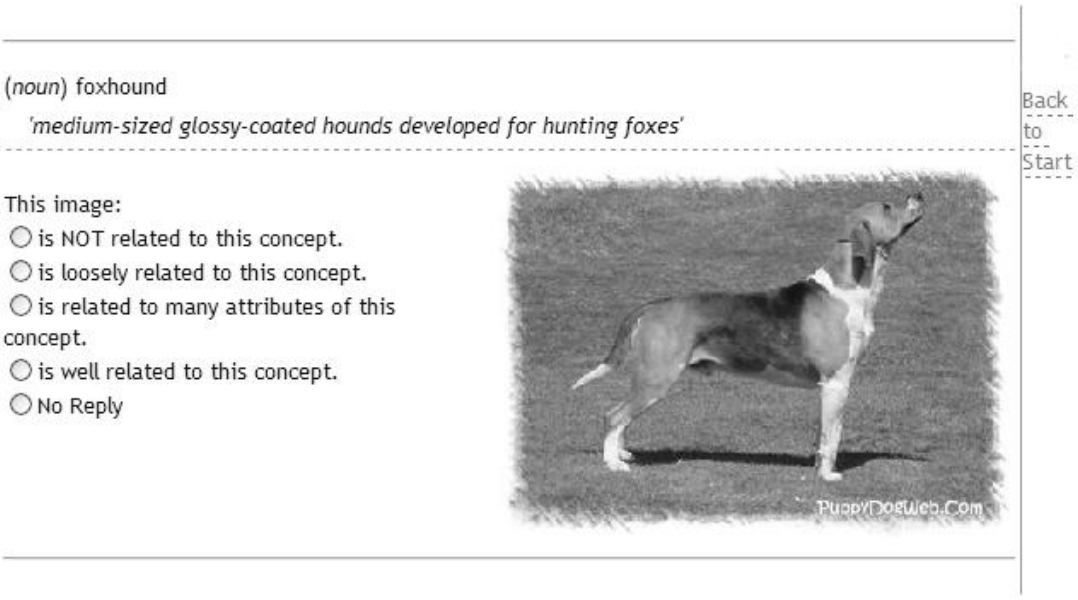



Figure 2: Image validation screen in PICNET

original links provided by WordNet and indirectly illustrate more difficult concepts by providing the natural associations humans use when trying to define or use such concepts. Given the availability of the images associated with such neighboring concepts, found by following semantic links in WordNet, users can "surf" by clicking on textual hyperlinks or image maps in the browser that bring them to spots in the network where they can add or validate concept-image associations.

Other possible activities include the use or integration of multi-media resources – such as animations, sound, or video, which could make an integral part of the user donated repository of concept illustrations. For instance, for certain categories of verbs, adjectives and nouns covering the auditory domain, it makes more sense to use sound samples rather than images.

## Ensuring Data Quality

Collecting from the general public holds the promise of providing much data at low cost. It also makes attending to an important aspect of data collection: ensuring contribution quality. We have implemented two main procedures to ensure the quality of the word/image associations collected through PICNET: (a) a scoring scheme that ranks word/image pairs based on total number of votes received from users of the various PICNET activities; and (b) administrative functions, which are used to monitor the quality of collected data.

### Scoring Synset/Image Associations

A complete history of users' decisions is maintained and used to rank the synset/image associations existing in PICNET. Each action provides an implicit quantified vote relating to the synset/image pair. The sum of these votes creates a score for the pair, allowing PICNET to dynamically rank images associated to a particular synset. The following list represents the possible actions that users can perform on the PICNET site, and the corresponding votes.

+5 Upload an image for a selected synset (from Search results)

+4 Image Validation (is well related to this concept)

+3 Image Validation (is related to many attributes of this concept)

+1 Image Validation (is loosely related to this concept)

-5 Image Validation (is NOT related to this concept)

+3 Free Association

+n Competitive Free Association (where n=number of players agreeing with the association)

### Administrative Functions

Since the goal of the project is to build the knowledge-base through the collaborative efforts of Web users, administrative functions must be kept to a minimum. However, an important concern with Web based data collection is the introduction of errors into the database because of user error or malicious intent. Additionally, there are strict rules regarding obscene images on governmentally-funded machines. These potential problems can be handled through some administration functions, which rely on the stored history of user activity.

All uploaded images must be verified by the administrator before they can enter the PICNET database. This is simply to judge that the image is suitable (i.e. family-friendly, adequate size and quality, not copyright restricted, etc.) for

use in the PICNET system. In fact, some of the these verifications, such as image size or quality, can be automatically performed using simple image processing libraries. Once the images have passed this validation step, the users receive their points for contribution, and the votes are scored.

Although this step may constitute a potential bottleneck, it is a necessary one. Oddly, the verification process is mildly compelling to a curious mind and can be performed at the average rate of 24 images per minute. Moreover, future versions of PICNET will consider giving administrator rights to selected PICNET users (similar to e.g. the collaborative DEX online project `http://dexonline.com`).

If malicious intent is discovered, the history mechanism makes it easy to rollback all of a particular user's activity.

### Preliminary Evaluations

Although at the time of writing PICNET was not publicly released, some preliminary evaluations concerning the quality of the results were conducted based on word/image associations collected from a small number of contributors (students and faculty in the Computer Science department at UNT). We conducted two different evaluations: (1) Average concurrence among users voting on the appropriateness of an image using the *Competitive Free Association* activity; (2) Quality of the top ranked word/image associations, based on the score computed using all the activities available in PICNET.

For the first evaluation, we measured the number of users voting for the same synset suggestion in each round, providing a measure of user concurrence. The average concurrence was around 43% with a standard variance of 0.05, which indicates a consistent agreement among three of five users contributing to this activity[3].

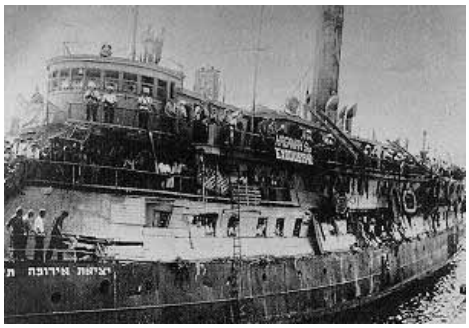
A manual verification of the top ranked word/image associations revealed the fact that the scoring scheme is successful in identifying good quality associations by combining the votes received from the Web-users contributing to the various PICNET activities. Figure 3 shows a sample of three word/image associations collected with PICNET.

## Related Projects

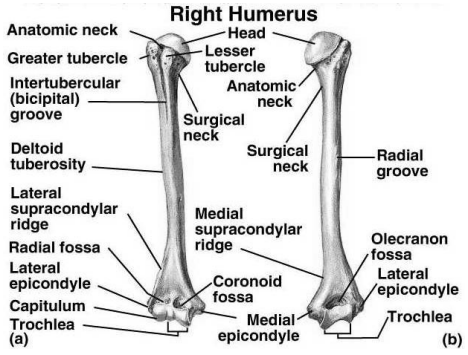
Collecting data over the Web for a variety of AI applications is a relatively new approach, which began with the broad *Open Mind* initiative (Stork 1999). The basic idea behind *Open Mind* (Stork 1999) is to use the information and knowledge obtainable from millions of Web users to create more intelligent applications. *Open Mind* projects related to natural language and world knowledge include the *Open Mind Word Expert* project (Chklovski & Mihalcea 2002), which builds lexically annotated corpora in a variety of languages using semantic labels collected from Web users, and the RSDNET project (Ayewah, Mihalcea, & Nastase 2003), which aims at building multilingual semantic networks by exploiting the knowledge of bilingual speakers. Other related projects include *Open Mind 1001*

[3]Each game consisted of five players. Three players are in concurrence, since two players (40%) have selected a third player's entry as the best available.

*exodus, hegira, hejira – (a journey by a large group to escape from a hostile environment)*



*humerus – (bone extending from the shoulder to the elbow)*




*Castro, Fidel Castro – (Cuban socialist leader who overthrew a dictator in 1959 and established a socialist state in Cuba (born in 1927))*

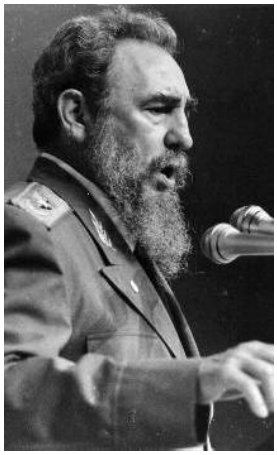


Figure 3: Sample of word/image associations contributed by PICNET USERS

*Questions* (Chklovski 2003) – a system for knowledge acquisition over the Web, and *Open Mind Common Sense* – a project (Singh 2002) that collects common sense statements from Web users.

Through the collaborative approach it implements, PICNET is most closely related to RSDNET, as it relies on similar techniques for contributing or validating information related to concepts defined in a semantic network. The main difference consists in the type of information that is sought: PICNET seeks pictorial representations for concepts in a dictionary to build richer illustrated semantic networks, while the goal of the RSDNET project was to collect textual representations of word meanings in a second language to the end of building bilingual semantic networks [4]. PICNET also relates to the ESPGame (von Ahn & Dabbish 2004) – a Web-based game that seeks to annotate images, although the main goal of PICNET is significantly different, as it seeks to associate images with word senses as defined in a dictionary.

## Conclusion

In this paper, we introduced PICNET, a Web-based system for building illustrated knowledge-bases by encoding word/image associations. We have shortly described a number of techniques to create an online pictorial semantic net that synergistically combines hyperlink based navigation with static illustrations of concepts, as well as dynamically generated interpolations for entities and relationships more difficult to visualize like abstractions or verbs. Another novel feature is reliance on images collected from and validated by online users, integrated with the use of automated image extraction techniques through image metasearch. Future work will focus on combining PICNET with educational tools and multi-lingual versions of WordNet - where common images will be used as a catalyst for user contributions in various languages.

## References


Ayewah, N.; Mihalcea, R.; and Nastase, V. 2003. Building multilingual semantic networks with non-expert contributions over the Web. In *Proceedings of the KCAP 2003 Workshop on Distributed and Collaborative Knowledge Capture*.

Barnard, K.; Johnson, M.; and Forsyth, D. 2003. Word sense disambiguation with pictures. In *Proceedings of the HLT-NAACL 2003 Workshop on Learning Word Meaning from Non-Linguistic Data*.

Chklovski, T., and Mihalcea, R. 2002. Building a sense tagged corpus with Open Mind Word Expert. In *Proceedings of the ACL 2002 Workshop on "Word Sense Disambiguation: Recent Successes and Future Directions"*.

Chklovski, T. 2003. *Using Analogy to Acquire Commonsense Knowledge from Human Contributors*. Ph.D. Dissertation, MIT.


[4] In fact, a significant fraction of the experience and code resulting from RSDNET was re-used in implementing the PICNET system


Fellbaum, C. 1998. *WordNet, An Electronic Lexical Database*. The MIT Press.

Mihalcea, R., and Tarau, P. 2004. TextRank – bringing order into texts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*.

Miller, G. 1995. Wordnet: A lexical database. *Communication of the ACM* 38(11):39–41.

Singh, P. 2002. The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access.*

Stork, D. 1999. The Open Mind initiative. *IEEE Expert Systems and Their Applications* 14(3):19–20.

von Ahn, L., and Dabbish, L. 2004. Labeling images with a computer game. In *Proceedings of the 2004 conference on Human factors in computing systems*.