TOWARD AN ONTOLOGY-ENHANCED INFORMATION FILTERING AGENT

Kwang Mong Sim

Department of Information Engineering, Chinese University of Hong Kong, Shatin, NT, Hong Kong.
Email: kmsim@ie.cuhk.edu.hk

Abstract— Whereas search engines assist users in locating initial information sources, often an overwhelmingly large number of *ULRs* is returned, and the task of browsing websites rests heavily on users. The contribution of this work is developing an *information filtering agent* (*IFA*) that assists users in identifying out-of-context web pages and rating the relevance of web pages. An *IFA* determines the relevance of web pages by adopting three heuristics: (i) detecting *evidence phrases* (*EP*) constructed from WORDNET's ontology, (ii) counting the *frequencies* of *EP* and (iii) considering the *nearness* among keywords. Favorable experimental results show that the *IFA*'s ratings of web pages are generally close to human ratings in many instances. The strength and weaknesses of the *IFA* are also discussed.

I. INTRODUCTION

Although search engines assist users in locating *URLs*, they often return an overwhelmingly large number of *URLs*, and the task of browsing websites rests heavily on users. This article presents the design of an *information filtering agent* (*IFA*) that extends the functionalities of search engines by assisting users in (i) filtering out-of-context web sites, and (ii) rating the relevance of web pages. An *IFA* filters irrelevant web pages by considering part-whole relations [1], and rates web pages by adopting three heuristics: (1) evidence phrases (*EP*) constructed from WORDNET's ontology [1], (2) *frequency* of *EP* and (3) *nearness* of keywords (section III). Section II discusses some of the guidelines adopted for ranking the relevance of evidence phrases such as synonyms, *hyponyms* (specialization) and *hypernyms* (generalization). Experimental results (section IV) show that an *IFA*'s ratings of web papges are generally close to human ratings in many instances. Section V discusses existing ontology-enhanced agents for information retrieval, and section VI summarizes the strength and weaknesses of the *IFA*.

II. RANKING RELATED INFORMATION

Since an *IFA* not only searches for exact keywords but also evidence phrases such as synonyms, *hyponyms* and *hypernyms* of keywords generated from WORDNET [1], this section establishes some guidelines for estimating the possible loss of information when a word is translated to its related term(s). The use of synonyms, hyponyms, and hypernyms have different effects on the query's *precision* and *recall* [2, pp. 248-249]. While precision refers to the proportion of the information retrieved that is relevant to a query, recall refers to proportion of relevant information retrieved. Although it is more apparent that a synonym is

closest in meaning to an original keyword C, its hyponym C_1 is assigned a higher precision score than its hypernym C_2 because a specialized term is *more likely* to fall within the scope of a user's interest than a generalized term. According to the definition of hyponym in [3, pp.38-39], C_1 contains the *meaning* of C; hence, C_1 has exactly the same features as C *plus* some additional ones. For instance, a user searching for websites on "domestic cats" is *more likely* to be interested in "Siamese cats" than "cats" in general (which may include wild cats such as cheetah and tiger). However, the recall of C_1 is *relatively lower* because the *extension* of C_1 (i.e., the set of entities referred to by C_1 [3,p.42]) is *included* in the extension of C_1. For example, a query on "Siamese cats" may not return documents on other types of cats such as "cheetah". The converse is true for C_2. The following heuristics are adopted by an *IFA* for rating the relevance of a *URL*.

1. A *URL* containing only an *exact word* is considered more relevant than a *URL* containing only its synonym.
2. A *URL* containing only the synonym of a term is considered more relevant than a *URL* containing only either its hyponym or hypernym.
3. A *URL* containing only the hyponym of a term is considered more relevant than a *URL* containing only its hypernym.

(1) deals with different contexts [3, p.38] in which words and synonyms are used. Even though in most situations, information about a concept and its synonym(s) is identical, there may be some exceptional cases. For instance, a chemistry student searching for web sites on "polysaccharide" (a synonym of sugar) may not necessarily find the advertisement on a special deal for "sugar" to be relevant for his chemistry course. The rationales of (2) are as follows: (i) translating a term to its hypernym may result in retrieving all documents on entities of the original term, plus some documents that may not be relevant (e.g., translating the query "domestic cat" to "cats" may retrieve documents about "cougar"), and (ii) translating a term to its hyponym may preserve precision, but may not retrieve some relevant documents. (3) is consistent with the explanations on C_1 and C_2 given above.

III. INFORMATION FILTERING AGENT

Filtering: The *IFA* filters out-of-context web pages that contain keywords in users' queries by considering *part-whole relations* of words such as *meronym* (part) and *holonym* (whole) [1] in WORDNET. Since users may input very short queries to search engine, there is little evidence to predict the relevance of a web page, and irrelevant web pages containing words with several meanings may be

among the suggested *URLs*. For example, a query for "battery" (in the sense of a "electric battery") may cause an out-of-context web site (e.g., http://www.geocities.com/~jmgould/batteryf.html) containing the word "battery" (but in the sense of a "gunnery") to be returned. The *IFA* identifies irrelevant *URLs* by searching for meronyms of query keywords of *other senses*. For instance, to search for irrelevant web pages for "battery" (in the sense of electric battery), the *IFA* filters web pages with meronyms such as "gun" and "missile launcher".

Rating ULRs: *URLs* that are *not* considered to be *irrelevant* will be rated by applying three heuristics based on [4].

Heuristic 1: The *IFA* searches for evidence phrases (**EP**) (such as exact keywords, synonyms, hyponyms and hypernyms) in a webpage *P* after removing all *function words* (e.g.,"and" and "of") [2, p. 279] in *P*. Exact matches, synonyms, hyponyms and hypernyms are assigned ratings of 1.0, 0.8, 0.6 and 0.4 respectively (see section IV). Algorithm 1 describes how *P* is rated based on **EP**.

Algorithm 1: Evidence Phrases
Let *Q* be the set of all words in a user query and *F* be the set of all function words in *Q*.
Let $\{k_1,...,k_n\}$ be the set of all keywords in $Q - F$.
Let *P* be a given web page and *W* be the set of words in *P*.
1.   Set **EP** = 0
2.   Remove all function words from *W*. Let the set of remaining words be $\{w_1,...,w_m\}$.
3.   For $k_i \in Q - F$
        If $k_i \in \{w_1,...,w_m\}$ then
                set **EP** = **EP** + 1
            else If synonym of $k_i \in \{w_1,...,w_m\}$ then
                set **EP** = **EP** + 0.8
            else If hyponym of $k_i \in \{w_1,...,w_m\}$ then
                set **EP** = **EP** + 0.6
            else If hypernym of $k_i \in \{w_1,...,w_m\}$ then
                set **EP** = **EP** + 0.4
          endIf
        endFor
4.   Set **EP** = **EP**/*n*

Heuristic 2: An *IFA* favors websites where evidence phrases occur more frequently. For instance, if "apples" occurred reasonably frequently in a web page *P*, it seems plausible to think that *P* contains information about "apples" [2, pp. 279-280]. An *IFA* determines the frequency of occurrence of evidence phrases (**EF**) of each keyword in a query *Q* in a website *U*. The frequency metric EF_L in a *URL* U_L is determined by the total number of exact matches, synonyms, hyponyms, and hypernyms of each of the *non-function* keywords of *Q*. Algorithm 2 determines the occurrence O_{ki} of *each* non-function keyword in a *URL* U_L. Unlike function words which seem to exhibit approximately almost equal frequencies of occurrence in all documents, non-function words occur

with greatly varying frequencies in different collections of texts [2, pp. 279]. Hence, it seems plausible to consider the frequency metric of the top 20 *URLs* with respect to non-function keywords in *Q*. The frequency of occurrence O_{ki} of each keyword in U_L is relative to that of the maximum frequency of occurrence O_{max-i} in a *URL* U_{max}. To normalize EF_L between 0 and 1, EF_L in U_L is computed by taking the average of *all* O_{ki}. O_{ki} of a non-function keyword k_i in U_L is determined by the ratio of (1) the total number of evidence phrases of k_i in U_L and (2) the maximum total number of occurrence O_{max-i} within the top 20 *URLs*. For example, consider the query "Renaissance painting", and U_L has 5 instances of "Renaissance" and 6 instances of "art" (hypernym of "painting"). Suppose the maximum frequencies of occurrences O_{max-1} and O_{max-2} of "Renaissance" and "painting" respectively are 25 and 50, then U_L receives a EF_L rating of ½ $[(O_{k1} / O_{max-1}) + (O_{k2} / O_{max-2})]$ = ½ $[(5 \times 1 / 25) + (6 \times 0.4 / 50)]$ = 0.248.

Algorithm 2: *Evidence Frequency*
1.   For the top 20 URLs = $\{U_1,...,U_{20}\}$ of *Q*, determine the frequency metric EF_L in U_L as follows:
    For all keywords $k_i \in Q - F$
        If k_i occurs in U_L then
                set $O_{ki} = O_{ki}$ + total number of k_i in U_L
        else
            If synonym of k_i occurs in U_L then
                set $O_{ki} = O_{ki}$ + 0.8 * total number of synonym of k_i in U_L
            else
                If hyponym of k_i occurs in U_L then
                    set $O_{ki} = O_{ki}$ + 0.6 * total number of hyponym of k_i in U_L
                else
                    If hypernym of k_i occurs in U_L then
                        set $O_{ki} = O_{ki}$ + 0.4 * total number of hypernym of k_i in U_L
                endIf
            endIf
        endIf
    endFor
2. For each keyword $k_i \in Q - F$, let U_{max-i} be a *URL* with the maximum number of occurrence O_{max-i} of k_i. Let *nf* be the number of keywords in $Q - F$. Compute EF_L in U_L as:

$$EF_L = \frac{1}{nf}\sum_{i=1}^{nf}\left[O_{ki} / O_{max-i}\right]$$

Heuristic 3: By considering *nearness* [2, pp. 236], the probable relevance of the information retrieved is likely to increase. If the query is "London symphony orchestra", and if "London", "symphony", and "orchestra" occur adjacently in a given web page *P*, then it is more likely that *P* contains more relevant information than when the words are separated by other words. In algorithm 3, the *IFA* (1) searches for clusters of words that match *exactly* those keywords in *Q* and (2) considers a cluster **C** that has the least number of keywords separating the first and last words. Since it is typical that users enter keywords with

no particular ordering (eg, users may submit queries in the form "Bus + Hong Kong"), this project has adopted the convention of searching for a cluster **C** with the *minimum* distance **D** regardless of the ordering of words in **C**. **D** is computed by considering the relative positions (indices) of the first and last word in **C**. If all words in **C** occur adjacently, **D** = 1, and $\mathcal{N}$ = 1. Note that $\mathcal{N}$ is normalized between 0 and 1.

Algorithm 3: *Nearness*
1. Find a cluster of words **C** = $\{c_1,...,c_n\}$ such that:
(i) $\{c_1,...,c_n\} \cap \{k_1,...,k_n\} = \{c_1,...,c_n\} \cup \{k_1,...,k_n\}$
(ii) the distance **D** between c_1 and c_n is minimum
2. Set **D** = [Index[c_n] – Index[c_1] + 1]/n
3. Set **nearness** $\mathcal{N}$ = 1/**D**

Combining the 3 heuristics: An *IFA* rates the relevance of a website as follows: **Relevance** = 0.34 * *EP* + 0.33 * *EF* + 0.33 * $\mathcal{N}$. Experimental results in section IV show that by placing almost equal weightings (0.34,0.33,0.33) for the 3 factors, ratings from a *IFA* coincide most with human ratings.

IV. EVALUATION AND RESULTS

Evaluation of the *IFA* consisted of (1) user study, and (2) a series of experiments using the *IFA* for rating the same set of web pages rated by human users. The agreements and differences between users' and agent's ratings were recorded and studied.

A. User Study

Query generation: Using 100 queries extracted from the MetaCrawler website http://www.metaspy.com/, two human users were asked to rate the relevance of the top 5 URLs returned by a search engine for these queries. Although space limitation precludes the 100 queries from being listed here, they were selected because they represent queries submitted by users to MetaCrawler.

Users' rating: Based in part on Ellis' [5] model of information seeking behaviors, each user is asked to perform both (1) browsing and (2) filtering the URLs.

(1) Browsing: Users were instructed to scan (rather than read or study) the content of a website for keywords and/or related information, and rate the relevance of the website by giving a rating for each of the following criteria:

(i) query words are found in the <u>title</u> or <u>headings of paragraphs</u>
• 0.2 if all query words are found
• 0.1 if some of the query words are found
• 0 if none of query words can be found
(ii) query words are found in the <u>first few paragraphs</u>
• 0.2 if all query words are found
• 0.1 if some of the query words are found
• 0 if none of query words can be found
(iii) <u>information related</u> to query words is found in the

first few paragraphs
• 0.2 if <u>related information</u> for *all* query words found
• 0.1 if <u>related information</u> for *some* of the query words found
• 0.1 if *no* <u>related information</u> can be found
(iv) query words are also found in the other parts of the web page, other than the <u>first few paragraphs</u>
• 0.2 if all query words are found
• 0.1 if some of the query words are found
• 0 if none of query words can be found
(v) <u>information related</u> to query words is found in other parts of the web page other than the <u>first few paragraphs</u>
• 0.2 if <u>related information</u> for *all* query words found
• 0.1 if <u>related information</u> for *some* of the query words is found
• 0.1 if *no* <u>related information</u> can be found

Criterion (i) is based on the rationale that users are often likely to browse the titles or sub-headings of information sources when looking for information on a particular topic. For instance, a person searching for monograph on "grid computing", typically looks at the title first [6] (e.g., "Grid computing: making the global infrastructure a reality"). Both criteria (ii) and (iii) check if the keywords or related terms appear at the beginning (eg., the first few paragraphs). The motivation is that a page relevant to the topic may likely mention those words or related terms right from the beginning [6]. Criterion (iii) is used to decrease the chance of rating favorably a web page that contains the search keyword but has a different meaning or usage context. For instance, if one searches for information about "battery" (in the sense of a "electric battery"), one should also be likely to expect to find information about "electrodes" and "voltaic battery", and "nickel-cadmium accumulators" rather than information about "artillery" or "battalion" (for instance, in the sense of a "gunnery"). As users browse the web page further, both criteria (iv) and (v) check for the presence of the search keywords or related terms. If both search keywords and related terms are found in the beginning and also other parts of the document, it seems intuitive to think that the web page should receive a higher score, provided that it is of the correct sense.

(2) Differentiating: If keywords and/or related terms are found, users are instructed to determine whether the information is of the correct sense. If the information is of the correct sense, the rating in (1) is retained. Otherwise, a rating of 0 is assigned.

User agreement: On average, the mean square error (*MSE*) between the ratings of the two users is small for most of the ULRs for the 100 queries. For over 90 of 100 queries, the average ratings for the top 5 URLs given by both users were within 10% *MSE* (Fig. 1). To minimize the possible subjective judgment of a single user, the average rating of both users is used for comparison with the rating of the *IFA*.
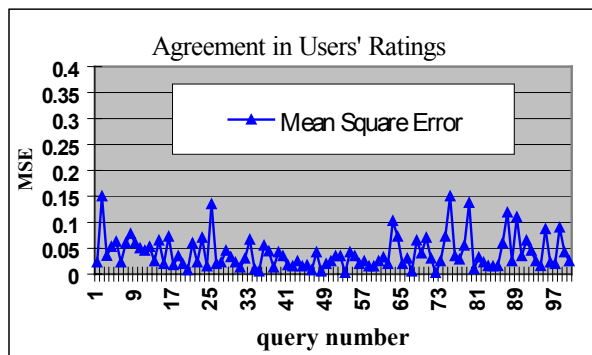
Fig. 1 Users' Ratings

B. Agent Rating

In this part, the *IFA* is programmed to rate the information content of the *same* set of 5 URLs for the *same* 100 queries used in the first part (user study). However, to show the effectiveness of enhancing the *IFA*'s performance using ontology, four sets of simulations were conducted. For the first set of simulations, the *IFA* was programmed to *incrementally* recognize exact words, synonyms, hyponyms, and hypernyms (see Table 1). Each of the four sets of simulations was repeated using ten difference weight patterns (Table 2). While section II discusses the ranking of the relevance of related information, the appropriate valuations of related terms such as synonym, hyponym and hypernym can only be determined through empirical studies. In evaluating the *IFA*, a total of 25 combinations of weight patterns and word relations were used. The rating of the *IFA* for a *URL* in each combination is compared with the corresponding average user rating for the same *URL*.

Performance Measure: The experiments were designed to compare both the ratings of human users and those of the *IFA* with regard to the information content of websites. This work considers that the *IFA* achieves good performance if the amount of difference between both human's and *IFA*'s ratings is small. Let R_I be the rating of the *IFA* for a *URL* and R_H be the average rating from the human subjects for the same *URL*. Let $SE = (R_I - R_H)^2$ be the square of error for each of the 5 *URLs*. For each query *Q*, the mean square error (*MSE*) is given as: $MSE = \frac{1}{5}\sum_{1}^{5} U_i$ such that $U_i \in \{U_1, ..., U_5\}$ is the set of 5 top *ULRs* of *Q*.

Empirical results: Two sets of results were obtained for investigating: (1) the appropriate valuation of related terms and (2) weighting of the three heuristics.

Appropriate valuation of related terms: Table 3 summarized the results obtained for *IFA1*, *IFA2*, *IFA3* and *IFA4* for the 10 weight patterns. Each element in Table 3 records the sum of *MSE* for the 100 queries for each combination of weight pattern and word relations. For instance, row 3 column 4 records the *MSE* for the 100

queries for *IFA4* (using exact words, synonyms, hyponyms and hypernyms) with a weight pattern of (1.0, 0.8, 0.6, 0.4). The following observations are drawn from Table 3:
(1) for the 10 weight patterns, (i) *IFA4* has the minimum average *MSE*, and (ii) the sum of *MSE* for *IFA3* is less than the average *MSE* for *IFA2*, which is less than the average *MSE* for *IFA1*.
(2) for *IFA1*, *IFA2*, *IFA3* and *IFA4*, using weight pattern 3, the *IFA* achieved the minimum sum of *MSE*. Among the weight patterns used, an *IFA* using (1.0, 0.8, 0.6, 0.4) have ratings closest to users'.

Weighting of the heuristics: Five combinations of weighting: *EP*, *EF* and $\mathcal{N}$ were tested. From the results shown in Table 4, the following observation can be drawn:
(3) among the weight patterns used, an *IFA* using pattern 3 has ratings closest to users'.

Table 1. IFA Simulation Sets

Simulation	Word relations
IFA1	{Exact words}
IFA2	{Exact words} + {synonyms}
IFA3	{Exact words, synonyms} + {hyponyms}
IFA4	{Exact words, synonyms, hyponyms} + {hypernyms}

Table 2. Sets of Weight Combinations

Weight Pattern	Exact Word	Synonym	Hyponym	Hypernym
1	1.0	0.9	0.8	0.7
2	1.0	0.9	0.7	0.5
3	1.0	0.8	0.6	0.4
4	1.0	0.9	0.8	0.5
5	1.0	0.5	0.3	0.2
6	1.0	0.7	0.3	0.2
7	1.0	0.6	0.4	0.2
8	1.0	0.3	0.2	0.1
9	1.0	0.8	0.6	0.2
10	1.0	0.4	0.3	0.2

Table 3. Differences between User and IFA Ratings

Weight Pattern	IFA1	IFA2	IFA3	IFA4	Average
1	17.4743	13.8222	12.6160	12.4392	13.9328
2	17.4743	13.8222	12.6694	12.4404	13.9269
3	17.4743	13.6487	12.5290	**12.2828**	13.7784
4	17.4743	13.8222	13.2389	12.8124	14.3065
5	17.4743	14.5109	13.3114	12.8778	14.5131
6	17.4743	14.4981	13.2771	12.7996	14.5152
7	17.4743	14.4555	13.2487	12.7963	14.4938
8	17.4743	14.5826	13.3284	12.8571	14.5606
9	17.4743	13.6487	12.5290	12.9906	14.1606
10	17.4743	14.5382	13.2986	12.8312	14.5356

C. Analysis and Discussion

From observation (1) in section IV.B, it can be concluded that by incrementally adding synonyms,

hyponyms, and hypernyms into the **EP** heuristic, the performance of the *IFA* in rating URLs is generally improved. This demonstrates the effectiveness of enhancing information filtering with ontological related keywords. For instance, for weight pattern 3, Fig. 2 shows that for many of the 100 queries, the average *MSE* when the *IFA* did not use ontological relation (*IFA1*) is considerably higher than when sub-class relations and synonyms are used (*IFA4*).

From observation (2), it can be concluded that among the weight patterns used, (1.0, 0.8, 0.6, 0.4) are generally appropriate valuations for exact word, synonym, hyponym and hypernym respectively. However, as there are infinitely many possible combinations of valuations, this work does not suggest that these valuations are optimal in the sense of achieving the minimum *MSE* between users and agent ratings.

Observation (3) showed that among the combinations used, by placing almost equal weighting on the three heuristics, an *IFA* achieved the minimum *MSE*.

Limitation and lesson learnt: Although observation (1) showed that by considering ontological relations of words, the *IFA*'s performance is enhanced, the improvement of the *average MSE* between *IFA1* and *IFA4* may appear to be less attractive (a modest 5.1% improvement). However, a closer examination (see Table 5) showed that, even though the average *MSE* improved by only 5.1%, there were about 40% of the queries in which the *MSE* of individual query improved by more than 5%. Moreover, it is noted that there were 24% of the queries with no improvement in the *IFA*'s performance. This is due to the limitation that no other evidence phrases can be generated for queries that contains keywords with no ontological relations such as names of persons (e.g, Tarkington) or places (e.g., Whitby). Additionally, it is also noted that there were 3 queries in which the *MSE* between users and *IFA4* (and *IFA1*) is quite large (above 30%) (see the 3 circled points in Fig. 2). This is because in its present form, the *IFA* is not designed to recognize images. For example, the *IFA* cannot recognize the pictorial images of words such "logo" and "Grand Hotel" that are well-understood by human users.

Part-whole relations: Even though meronymic relations are also useful sources of evidence phrases, unlike sub-class relations that share many properties (through inheritance), properties of concepts bounded by meronymic relations do not necessarily overlap. For example, websites containing the word "handle"(a meronym of "cup") may *not* be relevant to other queries, since "handle" is also a part of other unrelated artifacts such as "umbrella", "hammer" and "screw driver". For this reason, algorithm 1 in its present form does not include meronym and holonym in rating web pages. Nevertheless, preliminary investigation suggests that slight improvement in average performance may be achieved if the *IFA* also considers meronym when rating web pages. For instance, in a *URL* for the query "England Hotel", the (quite relevant) *URL* http://www.londonlodging.co.uk/ would otherwise receive a zero score from the *IFA*, because even though the word "London" (which is a part of "England") appeared several

times, the word "England" was not found in the web page. The main challenge to include meronym in the *IFA*'s rating mechanism is the problem of assigning an appropriate valuation. Current experiments adopt a valuation selected from the set {02,0.3,…,0.8}, and *preliminary* results seem to indicate that among these values, assigning a valuation of 0.3 to meronym records the relatively largest (albeit, modest) improvement in performance.

Table 4. Weighting of the 3 Heuristics

Weight pattern	*EP*	*EF*	$\mathcal{N}$	Difference between users and *IFA* ratings
1	0.6	0.2	0.2	29%
2	0.34	0.33	0.33	16%
3	0.5	0.25	0.25	21%
4	0.6	0.3	0.1	29%
5	0.4	0.4	0.3	18%

Table 5. Improvement of *MSE* for Individual Query

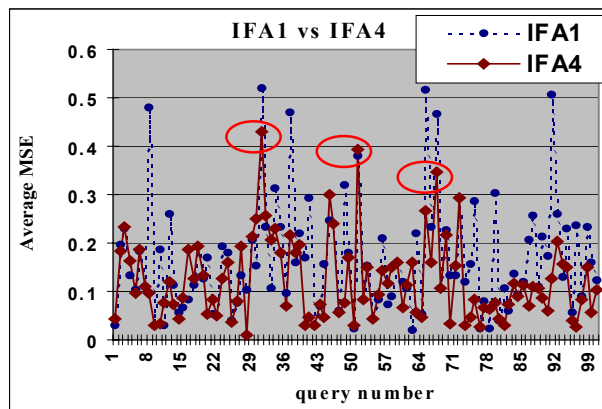| Improvement (*PI*)	No. of queries
20% ≤ *PI* < 40%	10
10% ≤ *PI* < 20%	15
5% ≤ *PI* < 10%	15
1% ≤ *PI* < 5%	36
PI ≈ 0%	24

Fig 2. Differences in User and Agent Ratings

V. RELATED WORK

Closest to this project is the work on *knowledge-enhanced search* [7]. Using background knowledge drawn from controlled vocabularies or ontological information sources, McGuinness's *FindUR* is a system that supports query expansion. *FindUR* improves recall and precision by searching for a set of related phrases constructed from super-subclass relationships, synonyms and instance relationships (instances of a concept are considered "evidence" for the concept). Another system that uses WORDNET is *Ontoseek* [8]. Designed for online yellow pages and product catalogues, *Ontoseek* is a content-based

information retrieval system. *Ontoseek* used *lexical conceptual graphs* to represent both queries and resource descriptions. In *Ontoseek* the problem of content matching reduces to ontology-driven graph matching. The *IFA* differs from *FindFur* and *Ontoseek*, since it also considers frequency and nearness of words. *DAML* (*DARPA* Agent Markup Language) [9] enriches the web with semantic contents to enable agents to understand what is on a web and to interact with the web. To resolve the problem that databases in different websites may use different words to refer to the same concept, *DAML* provides ontology pages (that define the relations among terms) on the Web. In *SHOE* (Simple HTML Ontology Extension) [10-11], HTML is augmented with semantics that allow users to encode useful ontological knowledge in web documents such as the classifications and relationships of concepts. SHOE enables WWW authors to annotate web pages with ontology-based knowledge that can be understood by *Exposε*, a web crawling agent in *SHOE*. *Exposε* can interpret *SHOE*-enabled HTML documents to augment their knowledge bases with information (such as "I am a graduate student") for answering queries (such as "Locate all graduate students in Florida working on machine learning") about these documents and their relationships. Although *DAML* is impressive and promising, it requires extensive restructuring and perhaps reengineering of the WWW. Before the full completion of the *semantic web* [9], one of the possible solutions is for agents to construct a set of evidence phrases for information filtering by consulting an ontology. Furthermore, if different ontology pages are used for different web services (different sites), the problem of using different terms for the same concept still arises in *DAML*. The *IFA* interacts with conventional HTML and consults only one ontology (WORDNET). Additionally, it is noted that *MELISA* (MEdical LIterature Search Agent) [12] is a prototype information retrieval agent that employs medical ontologies for reformulating and transforming a user's consultation into a collection of specific queries. *MELISA* allows different searches for a keyword using search modifiers. By consulting a medical ontology, it performs multiple queries for a term.

VI. CONCLUSION AND FUTURE WORK

An information filtering agent that assists users in filtering out-of-context webpages and assessing the relevance of webpages has been developed. Favorable results (section IV.A-B) show that the *IFA* is generally successful in rating the contents of web pages with a reasonably high degree of accuracy. However, as noted in section IV.C, the *IFA* has some limitations. In its present form, the *IFA* is not designed to recognize images. Hence, designing *IFAs* with both text-based and content-based filtering is among the list of agendas for future enhancement. Moreover, the problem of considering meronym in heuristic 1 is currently being investigated. The issue of designing *IFAs* that consider other WORDNET's ontological relations such as *coordinates* and *entailment* are currently being explored. Finally, it is reminded that the *IFA* is not designed to compete with existing search engines, but rather to extend and complement their functionalities. It is *not* directly intended to re-order the results returned by search engines, but perhaps to provide users with a tool for browsing websites on their behalf.

References

1. G.A. Miller. WORDNET: An On-line Lexical Database. International Journal of Lexicography 3-4, pages 235-312.
2. G. Salton. Automatic Text Processing. Addison Wesley, 1988.
3. F. Parker & K. Riley. Linguistics for Non-Linguists: A Primer with Exercises, (3rd ed.), Allyn & Bacon, 2000.
4. K.M. Sim. Web Agents with a Three-stage Information Filtering Approach. To appear in Proc. of the 2003 Int. Conf. on CYBERWORLDS, Dec. 2003, Singapore.
5. D. Ellis. Modeling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment. Journal of Documentation, 53(4):384-403, 1997.
6. D. Sullivan. How Search Engines Rank Web Pages. http://www.searchenginewatch.com/webmasters/rank.html.
7. D. L. McGuinness. Ontological Issues for Knowledge-Enhanced Search. In Proc. of Formal Ontology in Information Systems, pp. 302-316, 1998.
8. N. Guarino et. al. *OntoSeek: Content-Based Access to the Web*. IEEE Intelligent Systems, 14(3):70-80, 1999.
9. T. Berners-Lee et. al. *The Semantic Web*. In a feature article in Scientific American, May 2001.
10. S. Luke et. Al (1997). Ontology-based Web Agents. In Proc. of the 1st Int. Conf. on Autonomous Agents 1997, pp. 59-66.
11. J. Heflin et. Al (1999). Applying Ontology to the Web: A Case Study. In: J. Mira, J. Sanchez-Andres (Eds.), Proc. Int. Work-Conference on Artificial and Natural Neural Networks, Vol. II. Springer, Berlin, 1999, pp. 715-724.
12. J. Abasolo and M. Gómez. *MELISA*: *An ontology-based agent for information retrieval in medicine*. *ECDL* 2000 Workshop on the Semantic Web Lisbon, Portugal. Sep., 2000.