

Using a Word-Space Model to Determine the Relevance of Messages in Anchored Asynchronous Online Discussions *

Rodolfo Raga^a, Jennifer Raga^a, Erick Bonus^a, and Raymund Sison^b

^aComputer Science Department, Jose Rizal University,
80 Shaw Boulevard, Mandaluyong City, Philippines

^bCollege of Computer Studies, De La Salle University-Manila,
2401 Taft Avenue, Manila, Philippines

{jundy.raga, jennie.raga,erick.bonus}@gmail.com; sisonr@dlsu.edu.ph

Abstract. This paper presents results of the first phase of our study aimed at investigating the applicability of word-space models, particularly those generated using Random Indexing (RI) technique, for the task of determining the relevance of messages posted in anchored asynchronous online discussion forums. In this phase, we addressed several questions intended to establish baseline figures: How efficient will the word-space model perform in this task? How much of its classification decisions align with the decisions of the human annotators? More importantly, does the paradigmatic and syntagmatic contexts of words have a direct effect on its output and which produces the best results? Using Cohen's Kappa and Holsti's Coefficient Reliability measure, our experiments generated an initial reliability performance ($K=0.41$, $CR=0.72$). It further indicated that the syntagmatic context is more applicable for this task. We concluded with a discussion of the weaknesses identified and possible means of improving the level of performance.

Keywords: Asynchronous Online Discussion, Word-Space Model, Random Indexing, Paradigmatic Context, Syntagmatic Context.

1. Introduction

Asynchronous Online Discussions (AOD), also known as forums or threaded discussion boards, are a popular form of web-based computer-mediated communication (CMC). More recently, this form of communication has come into focus in education as an alternative form of assessment that can be used to supplement traditional classroom discussions. The advantage of this communication medium lies in its anytime and anywhere accessibility which provides logistic flexibility for students to communicate with each other and for teachers to assess their interaction.

A disadvantage of this medium however is that the focus of discussion is prone to topic drifting; this drifting is primarily due to the time-lags inherent in the asynchronous mode of discussion. Another problem is that the manual monitoring and assessment of message contributions are time-consuming and often tedious, especially for teachers. For these reasons, automated analysis for discussion understanding and better methods of topic-focus monitoring to enable more efficient information assessment are much sought for (Ravi & Kim, 2007).

In our study, we aim to investigate the feasibility of utilizing the functionality of word-space models for the task of analyzing online discussion transcripts. More particularly, we want to determine how word-space models can be used to detect the relevance of individual message posts

relative to the overall topic of discussion. Several methods for generating word-space models are available and have been successfully tested on detecting similarity between terms in groups of texts, but to our knowledge very few have been applied and tested on analyzing online discussion transcripts, a rare example is that of (McArthur and Bruza, 2003) which applied Latent Semantic Analysis on a small dataset of email correspondence to detect conversational implicatures. In our experiments, we chose to use the Random Indexing (RI) technique. RI is a word-space modeling technique primarily designed to measure similarity between terms and not between documents. In our work, we extended the capability of RI to work both at the document and message size levels.

The rest of the paper has the following organization: The second section endeavored to put our study into context by providing a brief background on the problem of topic drifting in online discussions and cites a recent approach used to address it. The third section discussed the concept of word-space models and provides a brief introduction on how it can be implemented using RI. The fourth section presented a description of the experiments we conducted. In the fifth section we presented the results of the experiments and provided some discussion. We ended the paper with our conclusion and expected future directions.

2. Review of Literature

The topic focus of online discussions constantly changes from one topic to another (Beaudin, 1999). This phenomenon is what (Potter, 2007) referred to as *topic drifts* or the tendency of online discussions to stray from their announced topic.

In academic settings, researchers try to limit the occurrence of this problem by implementing anchored discussions (Guzdial and Turns, 2000). These educational online discussions are structured to align the contributions of participants to a single topic represented by the contents of a reference document called the *anchor document*. Literature shows that this method yields a more coherent discussion than traditional forums by promoting conscious topic monitoring (van der Pol et al, 2006). However, participants are unconsciously still prone to introduce irrelevant posts which often induce off-topic discussion. As such, tools that can help mediators automatically detect, at the earliest, such irrelevant contributions are still needed.

An initial requirement for the development of such tools is a method for determining the alignment of each message to the overall topic focus of the discussion. This necessitates a content analysis approach similar to the strategy proposed by (Teufel and Moens, 1998) whereby the alignment of one text with another was measured by a mechanism that analyzes several characteristics of the aligned sentences such as the presence of particular phrases and occurrence of thematic words and proper names. For our purpose, we are looking towards utilizing the functionalities of word-space models for the same task. The context of the work we're pursuing is the anchored discussion mentioned above. As such, the basis of the relevance value we are trying to measure is the closeness of the semantic information between the contents of each message to the contents of an anchor document.

Word-space models have previously been observed to provide two modes of semantic similarity between terms. Sahlgren (2006) noted that these modes depend on whether the paradigmatic or syntagmatic relationship of the terms within a context are captured. Syntagmatic context refers to the linear relationship of words and applies to linguistic entities that occur in sequential combinations while paradigmatic context refers to the substitutional relationship of words and applies to words that can be used in the same context but not at the same time. We aim to initially investigate, among other things, whether or not these two modes will provide different levels of performance for this task, and if so, which provides the best result.

3. Word-Space Models and Random Indexing (RI)

In this section, we present a simple introduction to the concept of word-space modeling, how it is used to assign meaning to a term, and how it can be implemented using Random Indexing.

3.1. Word Space Model

A word-space model is a spatial representation that derives the meaning of words by plotting these words in an n -dimensional geometric space (Sahlgren, 2005). This process is similar to the way points are plotted in a two dimensional graphing paper. The main difference is that, in the case of a word-space, the dimension n can be arbitrarily large. The size of this dimension is determined by the number of unique word type in the set of words to be plotted.

Usually, the coordinates used to plot each word depends upon the frequency of the contextual feature that each word co-occur with within a text. For example, words that do not co-occur with the word to be plotted within a given context are assigned a coordinate value of zero. The set of zero and non-zero values corresponding to the coordinates of a word in a word-space are recorded in a *context vector*. Because most of the words in any text dataset will never co-occur with a particular word, the context vectors are often sparse or full of zero values.

By itself, the position of a word in a word space does not indicate anything about its meaning. To deduce a certain level of meaning, this position needs to be measured relative to the position of other words. In this sense, a linguistic concept known as the *distributional hypothesis* which states that “*words that occur in the same contexts tend to have similar meanings*” is applied. Having similar contexts means that words are surrounded or that they co-occur with same set of words. Thus, if we plot these words in a word-space they would be positioned close to each other. The level of closeness of words in the word-space is often referred to as the spatial proximity of words. This spatial proximity is what is used to represent the semantic similarity of words. A common approach used to determine spatial proximity is to measure the cosine of the angle generated between the plotted context vectors. The formula for computing cosine is as follows:

$$CoSim(Q,D) = \frac{Q * D}{|Q| * |D|} \quad (1)$$

Where: Q is a vector representing one term or document, D is a vector representing another term or document related to Q , and $|Q|$ and $|D|$ are the magnitudes of Q and D .

Currently, there are three major approaches to implement a word space. These include: Latent Semantic Analysis (LSA), Hyperspace Analogue to Language (HAL), and Random Indexing (Sahlgren, 2005). For our purpose, we opted to use the Random Indexing approach.

3.2. Random Indexing

Random Indexing (RI) is a word space modeling technique that can be used with any type of linguistic context. It is inherently incremental and incorporates a built-in dimension reduction phase (Sahlgren, 2005).

There are two basic steps involved in using Random Indexing to implement a word-space:

1. The first step involves the assignment of a unique and randomly generated label called an *index vector* to each word in the data. These vectors are sparse, high-dimensional, and ternary. High-dimensional means that it uses a large number of dimensions while ternary means that the label consists of a small number of randomly distributed +1s and -1s, with the rest of the elements of the vectors set to 0.
2. Then, context vectors with the same number of dimensionality as the index vectors are automatically produced by scanning through the text, and each time a word occurs in a context (e.g. in a document, or within a sliding context window), that word's index vector is added to the context vector for the word in question.¹

RI has attracted much attention with successful applications in term similarity measurement. For our purpose, we have selected RI because of its incremental approach; this means that the

¹ Please refer to (Sahlgren,2005) for more details

context vectors can be used for similarity computations even after just a few examples have been encountered. This characteristic is useful in experiments and applications such as ours where the size of the dataset is small. Secondly, as already mentioned, Random Indexing can be used with any type of context. This characteristic is useful in our experiment because we want to separately capture the paradigmatic and the syntagmatic contexts of words.

4. Experiments

4.1. Research Questions

The experiments described below are meant to address the following questions:

1. Using only the contents of an external anchor document as basis, how efficient is the word-space model's baseline performance in identifying relevant and irrelevant messages in an online discussion transcript?
2. To what extent will the classification decisions generated by the word-space models for this task align with the classification decisions made by the human annotators?
3. Does the paradigmatic and syntagmatic relationship of words have a direct effect on the performance of the word space model for this task? If so, which of the two provides the best performance?

4.2. Model Implementations

To provide an answer to the above enumerated questions, we need to separately capture the syntagmatic and paradigmatic relationships of words and use it to build a word-space. This entailed implementing two variants of context windows that dissected the sentences in each document into its syntagmatic and paradigmatic context as described below:

4.2.1. Syntagmatic Context Window

Following the definition specified in (Sahlgren, 2006), we defined the syntagmatic context vector v as constituting of n context regions c as follows:

$$\vec{v} = (c_1, c_2, \dots, c_n) \quad (2)$$

We then defined each context region c as a window that we deem would capture the syntagmatic context of words in a sentence as follows:

$$c = (t_1, t_2, \dots, t_m) \quad (3)$$

where c is a single context, $t_1..t_m$ are m terms occurring in sequential order within the text being modelled and m is a number defining the size of the window. This window slides through each sentence in the text, one word at a time. As such, some windows will necessarily have null values in them (those that are near the end of the sentence).

In RI terminology, the value we assign to each context c is the sum of the index vector of m terms occurring sequentially in it. For weighing function, we multiplied the index vector of each term to its position index within the context (i.e. the 1st term by 1, the 2nd by 2, and so on).

4.2.2. Paradigmatic Context Window

Similarly, following the definition specified in (Sahlgren, 2006), we defined the paradigmatic context vector v as constituting of n word types fw as follows:

$$\vec{v} = (fw_1, fw_2, \dots, fw_n) \quad (4)$$

The value we assigned to each word fw is determined by its surrounding words as contained in a window that we deem would capture the paradigmatic context of the word as follows:

$$w = (wp_m, wp_{m-1}, \dots, wp_1, fw, ws_1, ws_2, \dots, ws_m) \quad (5)$$

here, fw is the focus word and $wp_m, wp_{m-1}, \dots, wp_1$ and ws_1, ws_2, \dots, ws_m are groups of m sequential words that precede and succeed the focus word respectively within the text to be modeled. This window is also a sliding window and rolls through each sentence in the text, one word at a time.

In RI terminology, the value we assign to each focus word fw is the sum of the index vectors of m sequential terms surrounding fw in some $m + m$ context window. For weighing function, we multiplied the value of each surrounding word to their distance from the focus word.

4.3. The Discussion Transcript

The discussion transcript we used as dataset in our experiments was culled from a public forum. We deemed this transcript as suitable for our purpose since the discussion that generated it revolved around an online article which was cited in the first message. The online article consists of 34 sentences with a total of 602 words and served as the *anchor document*. Table 4 provides some statistical details on the discussion transcript.

Table 1. Some statistical details of the discussion transcript.

Total number of messages posted	87
Total number of participants	16
Discussion duration	5 days
Average message length	86.3 words

4.4. Methods

The methods we used in this study involved four steps requiring as input a text-based transcript of an online discussion and an external anchor document. The output included a relevance measure for each message in the transcript as well as calculations of reliability statistics.

4.4.1. Initial Setup

First, the text of the selected discussion transcript was downloaded along with the online article that served as the *anchor document*. These were then encoded into separate tables in a database. This task was accomplished using the Microsoft Access software.

We then asked three people to annotate the messages in the discussion transcript. The annotators were instructed to read first the anchor document, after which they were asked to give a binary judgment as to whether they think each message is relevant or irrelevant to the topic expressed in the document {1 = relevant, 0 = irrelevant}. The annotators performed this task independently, without convening with each other. The technical background of the annotators also vary; one is a computer programmer that had extensive hands-on experience with the software tool discussed in the anchor document, one is a computer science instructor having theoretical understanding but little industry-based experience, and the last is a call-center agent who has no background knowledge nor experience in programming.

4.4.2. Dataset Preprocessing

Second, a tool was constructed to browse through each message body in the database as well as through the anchor document to perform dataset pre-processing. Pre-processing included replacing special symbols and numeric characters with null values, rolling upper-case alphabetic

characters into lower case, and segmenting each text into its sentential units. The delimiters used to identify sentential units include the symbols [., ?, !, :].

No stopping or stemming of words was applied to the text. However, within the RI computations, the random index of stopwords was set to zero. This has the same effect as removing those words from the text. The stoplist we used was culled from (Sanderson, n.d.).

4.4.3. Experimental Setup

In our experiments we first divided the discussion transcript into two distinct dataset. The first dataset included messages that were unanimously classified by the annotators as being either relevant or irrelevant; this dataset served as our Gold Standard and we refer to it as the *unanimous dataset* (n=66). The annotators unanimously coded 49 messages as relevant and 17 messages as irrelevant. The other dataset included all the messages; we refer to this dataset as the *standard dataset* (n=87).

We then conducted two sets of experiments: one using the paradigmatic context and the other the syntagmatic context. In both cases, the RI system was first trained on the anchor document before being applied to the datasets. Also, the generated word space was first tested on the *unanimous dataset* before applying it to the *standard dataset*. The purpose of the first test is to establish a baseline on the performance of the word space model in identifying the exact category of a message. The second test was meant to probe the reliability of the decisions of the model as compared to humans with all the noise present.

We also experimented on various window sizes. For the syntagmatic context the window sizes we used are $w = \{2, 4, 6, 8\}$ while the paradigmatic counterpart are $w = \{1, 2, 3, 4\}$. The Random Indexing system we used in our experiments was constructed using the JavaSDM package (Hassel, 2004) and the parameters we applied are shown in Table 2:

Table 2: Parameters used in the RI implementation

RI Parameters	Values Used
Number of Dimensions used	900
Degree of Randomness Used	8
Seed-value used to generate random number	123

4.4.4. Reliability Measures

In this final step, we first measured the precision and recall of the classification decisions of the model in classifying messages in the *unanimous dataset*. Then, reliability measures were taken comparing the classification decisions of the human annotators with those produced by the word space model in the *standard dataset*. Two reliability measures were employed for this purpose: Holsti's coefficient of reliability (*CR*) which measures the agreement between two annotators divided by the total number of messages analyzed and Cohen's kappa (*K*) which computes the proportion of agreement actually observed between annotators after adjusting for the proportion of agreement expected by chance².

To compare the RI system's classification decisions with those of our human annotators, we treated any message with a computed cosine between 0.1 and 1.0 to be relevant (category = 1). Otherwise, it was treated as not-relevant (category = 0). Processing and computations in this final stage was done using the Microsoft Excel software.

5. Results and Discussions

² Acceptable values for (*K*) are between 6.0 – 7.0 while (*CR*) values can range from 0.00 (indicating no agreement) to 1.00 (indicating complete agreement).

5.1. Classification Precision and Recall

Tables 3 and 4 below shows the precision and recall gathered using the *unanimous dataset*.

Table 3: Precision and Recall of the classification results using the syntagmatic context

	$w = 2$	$w = 4$	$w = 6$	$w = 8$
P-R	0.57	0.57	0.55	0.52
R-R	0.71	0.71	0.71	0.65
P-IR	0.89	0.89	0.89	0.87
R-IR	0.82	0.82	0.80	0.80

w = window size, P-R = precision on relevant messages, R-R = recall on relevant messages, P-IR = precision on irrelevant messages, R-IR = recall on irrelevant messages.

Table 4: Precision and Recall of the classification results using the paradigmatic context

	$w = 2$	$w = 4$	$w = 6$	$w = 8$
P-R	0.30	0.31	0.31	0.31
R-R	0.47	0.47	0.47	0.47
P-IR	0.77	0.78	0.78	0.78
R-IR	0.61	0.63	0.63	0.63

w = window size, P-R = precision on relevant messages, R-R = recall on relevant messages, P-IR = precision on irrelevant messages, R-IR = recall on irrelevant messages.

5.2. Reliability

Tables 5 and 6 below present the reliability measures gathered using the *standard dataset*.

Table 5: Average Kappa and Coefficient Reliability measures with the human annotators using the syntagmatic context.

	$w=2$	$w=4$	$w=6$	$w=8$
Ave. Kappa	0.41	0.37	0.33	0.30
Ave. CR	0.72	0.71	0.69	0.69

Table 6: Average Kappa and Coefficient Reliability measures with the human annotators using the paradigmatic context.

	$w=2$	$w=4$	$w=6$	$w=8$
Ave. Kappa	0.12	0.14	0.14	0.14
Ave. CR	0.57	0.58	0.58	0.58

5.3. Discussion

Initial interpretations we derived from our experiments are three-fold, relative to the three questions presented in section 4.1. In our analysis, we divided the messages into four groups: (1) those that were correctly classified as relevant (true positive), (2) those that were correctly classified as irrelevant (true negative), (3) those that were incorrectly classified as relevant (false positive), and (4) those that were incorrectly classified as irrelevant (false negative).

Question #1: Using only the contents of an external anchor document as basis, how efficient is the word-space model's baseline performance in identifying relevant and irrelevant messages in an online discussion transcript?

As shown in tables 3 and 4, the highest precision-recall tandem for classifying relevant messages (P=0.57, R=0.71) as well as classifying irrelevant messages (P=0.89, R=0.82) from the *unanimous dataset* was returned by the RI system implementing the syntagmatic context and using a window size of 2 and 4. The paradigmatic counterpart of these settings returned lower values both for relevant (P=0.31, R=0.47) and irrelevant messages (P=0.78, R=0.63).

These results seem to indicate that the word-space model is more efficient at identifying irrelevant messages for this situation. To establish some basis for this performance, we manually analyzed the results on the syntagmatic context. We found that messages that were categorically assigned to the *true positive* group all contain several technical keywords that also occurred in the anchor document; we initially assumed that the presence of these keywords is the primary reason why these messages were classified as relevant. However, we also found that the *true negative* group is not exclusive to the messages that contained no shared technical keywords; some messages in this group also contained technical keywords that occurred in the anchor document, one of them even contained as much as three keywords. This finding seemed counter-intuitive with the first assumption we made; it implies that while the model relies on the presence of technical keywords to identify relevant messages, it doesn't necessarily rely on the absence of the same to identify irrelevant messages.

Further analysis led us to consider the way technical keywords are used in each message as a probable cause of this phenomenon. We noticed that the anchor document, being a technical document utilized and prescribed technical keywords in proximity with each other. This characteristic can also be observed in most messages that contained technical discussions. As such, when compared with the anchor document, technical messages got a higher chance of being classified as relevant. All messages assigned to the *true positive* group are of this type.

On the other hand, most non-technical messages are loosely constructed. Some of these messages also cited technical keywords but often presented them either in wide intervals or in isolation. Thus, when compared with the compact structure of the anchor document, the system has a higher tendency to tag them as irrelevant even though they contained technical keywords.

What we deduced from this observation is that the word-space model may be more efficient at identifying irrelevant messages in this situation because it can utilize two sources of information: the first is the absence of technical keywords, and, if technical keywords are presented, it can analyze the proximity of these keywords to determine whether they project the same topic as those of the anchor document. The analysis on relevant messages, on the other hand, solely relied on the presence of technical keywords. It has no alternative for recognizing the relevance of messages that have no keywords present. This problem is further compounded by the presence of dominant words (i.e., in our dataset, the word "sqldatasource") in the messages. Messages assigned to the false positive group seemed to indicate that the model have a tendency to classify as relevant those messages that contained the most frequently occurring word in the anchor document even if this word occurred in isolation.

Question #2: To what extent will the classification decisions generated by the word-space models for this task align with the classification decisions made by the human annotators?

As indicated in the values presented in Tables 5 and 6, both the paradigmatic and syntagmatic implementations of RI produced unacceptable values of reliability at this phase. With the RI system implementing the syntagmatic context and using a window size of ($w=2$) producing the higher reliability score (K=0.41, CR=0.72) with the human annotators. Manual inspection of the messages assigned to the false positive and false negative groups revealed that most of the discrepancies incurred on these tests were made on messages that either expounded on a particular concept or led the discussion to an off-topic direction. In the case of the latter, participants introduced other technical keywords in their messages that are related to but are not found in the anchor document. We assumed that this caused discrepancy because the RI system

has no way of recognizing these keywords, especially if they are embedded within analogies. In the former, messages contained questions or brief comments that presented isolated technical keywords. If the technical keyword used is dominant (i.e., the highest occurring), the RI system classified the message as relevant otherwise it is classified as irrelevant. Human annotators, however, are still able to methodically discriminate the true value of these messages. We surmised that unconsciously they do this by utilizing the coherence or semantic relatedness of successive messages within the transcript.

At any rate, we found that the best performance of the model aligned more closely to the decisions made by the computer science instructor-annotator. Table 7 provides the details of this observation. Although this finding is still open to interpretations, we deem that this gives a good background and motivation for the application that we are envisioning.

Table 7: Best performance of the word-space model using syntagmatic context ($w=2$)

	Programmer	Instructor	Call Center Agent
Kappa	0.41	0.44	0.37
CR	0.72	0.74	0.70

Question #3: Does the paradigmatic and syntagmatic relationship of words have a direct effect on the performance of the word space model for this task? If so, which between the two provides the best performance?

Clearly, the results presented in Tables 3, 4, 5, and 6 indicated that context has a direct influence on the performance of the model. Another interesting observation that can be deduced from these results is that the RI implementing the syntagmatic context outperformed the RI implementing the paradigmatic context. Two related and supporting hypotheses may be derived from this discrepancy. One, is that the syntagmatic context, for this task, are better geared to model the human annotators' classification decisions than the paradigmatic context, the other is that human annotators may be relying more on the syntagmatic context of words in determining the relevance of messages under the circumstances given. Further studies are needed to prove these hypotheses.

By intuition, we believe that since information coming from both forms of contexts is available to them, the human annotators are utilizing both of these in finalizing their classification decisions (i.e., the arrangement of the words + the acquired meaning of each word). We argue that this, to a greater extent, caused the misalignment between the decisions exhibited by the word-space model and those made by the human annotators in our experiments.

6. Conclusions and Future Directions

Analyzing the applicability of the word-space model for the task of determining the relevance of forum messages, undoubtedly, requires a multi-phase process. In this paper, we have started exploring the capabilities of the model and gathered figures that will serve as our baseline data. Initially, we have shown the insufficiency of the word-space model's functionality to produce reliable results given the conditions specified. In succeeding phases of our study, we aim to explore other means that may improve the performance level of the word-space model. Some of our prospects include the following:

1. Other datasets. It would be interesting to experiment with other datasets, preferably, on larger discussion transcripts but of similar anchor document sizes. This will enable us to determine whether the observations cited in answering question #1 represented recurring patterns or merely an artifact of the dataset we used.
2. Word sense information. In the current model, we tested the performance of the syntagmatic and paradigmatic contexts separately. In future experiments we aim to find ways of combining the functionalities of the two contexts. For example, the paradigmatic

context seems more applicable to determining word similarity or antonymity. It would be interesting to determine whether this context can serve as an effective source of word-sense information for the syntagmatic context. This may help improve the precision and recall of the word-space model's performance.

3. Message hierarchy information. In the current model, we treated each message as an independent document. However, in reality, messages in a discussion transcript are inter-related, with previous messages possibly providing a context to succeeding messages. Humans implicitly recognize this relationship but our RI system cannot. We aim to also take this into consideration in future experiments and to enable the system to adjust the relevance scores based on the contextual relationship of each message to its preceding messages. This may help improve the reliability measure.

The rising popularity of Asynchronous Online Discussion as a tool for supporting student learning necessitates the development of tools that can be used to efficiently monitor and assess online discussions. Word-space models provide an alternative approach for developing these tools. Current applications of word-space models focus on extracting meaning using large-scale corpora, more so, traditional tests of how well each particular model represents meaning largely revolves around word level comparison. The contribution of this paper is that it provides an initial bench-mark on the performance of this approach on small-sized datasets but on larger linguistic structures. We believe that such benchmarks are vital for fine-tuning word-space models for this particular task and application.

References

- Beaudin, N.B. 1999. Keeping Online Asynchronous Discussions on Topic. *Journal of Asynchronous Learning Networks*, Volume 3, Issue 2 - November 1999
- Guzdial, M., and J. Turns. 2000. Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences*, 9, 4, 437-470.
- Hassel, M. 2004. JavaSDM - A Java package for working with Random Indexing and Granska. <http://www.nada.kth.se/~xmartin/java/JavaSDM/>
- McArthur, R., and P.D. Bruza. 2003. Discovery of tacit knowledge and topical ebbs and flows within the utterances of online community, in: *Chance Discovery*, Y. Ohsawa and P. McBurney eds., Springer Verlag, pp. 115–132.
- Potter, A. 2007. *An Investigation of Interactional Coherence in Asynchronous Learning Environments*. Ph.D. thesis, Nova Southeastern University
- Ravi, S. and J. Kim. 2007. Profiling Student Interactions in Threaded Discussions with Speech Act Classifiers. In *Proceedings of Artificial Intelligence in Education Conference, 2007*.
- Sahlgren, M. 2006. The Word-Space Model. Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high dimensional vector spaces. Ph.D. thesis, Stockholm University.
- Sahlgren, M. 2005. An Introduction to Random Indexing. In *Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering*, Copenhagen, Denmark, 2005.
- Sanderson, M. (n.d.) IR linguistic utilities – Stop word list. Retrieved on June 1, 2008 from http://www.dcs.gla.ac.uk/idom/ir_resources/linguistic_utils/stop_words
- Teufel, S., and M. Moens. 1988. Sentence extraction and rhetorical classification for flexible abstracts. In *AAAI Spring Symposium on Intelligent Text Summarization*, Stanford.
- van der Pol, J., W. Admiraal and P.R.J. Simons. 2006. The affordance of anchored discussion for the collaborative processing of academic texts. *Computer-Supported Collaborative Learning, 2006*. 1: 339–357