

# Genetic Algorithm for Analysis of Abdominal Aortic Aneurysms in Radiology Reports

Robert M. Patton  
Oak Ridge National  
Laboratory  
P.O. Box 2008 MS 6085  
Oak Ridge, TN 37831  
1-865-576-3832

pattonrm@ornl.gov

Barbara G. Beckerman  
Oak Ridge National  
Laboratory  
P.O. Box 2008 MS 6085  
Oak Ridge, TN 37831  
1-865-576-2681

beckermanbg@ornl.gov

Thomas E. Potok  
Oak Ridge National  
Laboratory  
P.O. Box 2008 MS 6085  
Oak Ridge, TN 37831  
1-865-574-0834

potokte@ornl.gov

Jim N. Treadwell  
Oak Ridge National  
Laboratory  
P.O. Box 2008 MS 6085  
Oak Ridge, TN 37831  
1-865-574-5530

treadwelljn@ornl.gov

## ABSTRACT

An abdominal aortic aneurysm is a problem in which the wall of the artery that supplies blood to the abdomen and lower extremities expands under pressure or balloons outward. Patients must undergo surgery to repair such aneurysm, and there is currently no known indicator of success or failure from this surgery. Our work uses a genetic algorithm to analyze radiology reports from these patients to look for common patterns in the language used as well as common features of both successful and unsuccessful surgeries. The results of the genetic algorithm show that patients with complications or unusual characteristics can be identified from a set of radiology reports without the use of search keywords, clustering, categorization, or ontology. This allows medical researchers to search and identify interesting patient records without the need for explicitly defining what “interesting” patient records are.

## Categories and Subject Descriptors

I.2.6 [Learning]: analogies, concept learning, connectionism and neural nets, induction, knowledge acquisition, language acquisition, parameter learning.

H.3.3 [Information Search and Retrieval]: clustering, information filtering, query formulation, relevance feedback, retrieval models, search process, selection process.

## General Terms

Algorithms, Design

## Keywords

Genetic algorithm, abdominal aortic aneurysm, medical knowledge discovery, natural language processing

## 1. INTRODUCTION

Genetic algorithms have been used in medical applications to characterize features and findings identified in reports. Previous

work has been done on the application of such algorithms to mammography reports to classify the data [6][7], but it is important to test these concepts in broader applications to determine scalability and appropriateness across a spectrum of patient data. Ultimately, the use of GAs could become standard for data analytics in large medical data warehouses. To test the hypotheses that a GA developed for mammography could be used with new data types, it has been extended for use with patient’s having Abdominal Aortic Aneurysms (AAAs).

An AAA is a problem in which the wall of the artery that supplies blood to the abdomen and lower extremities expands under pressure or balloons outward. In most cases, there are no advance symptoms, and patients only learn they have an aneurysm when they are admitted to the hospital for other problems for which an abdominal computed tomography (CT) scan is ordered. Aneurysms in danger of rupture are repaired surgically by the insertion of a graft or stent. Then, the patient is followed for several years to ensure there are no further problems such as secondary ruptures, slippage or leakage (endoleaks). Characterization and pattern analysis of these secondary problems has never been done using automated tools for extraction, comparison, or temporal comparison. Generally, the surgeon relies on the CT scans, radiology reports, and states of the patient’s health during routine monitoring to assess if a repair is needed, but in most cases, these problems are only detected during the follow up scans or visits to the emergency room. A far more effective way to support clinical decision making would be to employ tools such as GAs in an integrated learning environment to assess all the contents of the radiology reports, operative notes and other supporting patient information. This environment would allow for more complete characterization of AAA features, changes, types of aneurysm, and assessment of patient care. For example, there are four types of leaks that may occur after the repair of an AAA. One of these is life threatening and requires immediate surgery. Another, the most common type, frequently repairs itself or goes away over time, and may not need surgical intervention. A third type deals with the type of graphs, and the fourth is rarely seen. If surgeons could predict, through analysis of the reports and images available to them when a leak or rupture is likely to occur and the type of leak, they could be in a better position to help their patients. The challenge lies in developing or applying tools that can accurately assess and organize the language of the patient reports.

Copyright 2010 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the U.S. Government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy.

GECCO’10, July 7–11, 2009, Portland, Oregon, USA.  
Copyright 2010 ACM 978-1-60558-505-5/09/07...\$5.00.

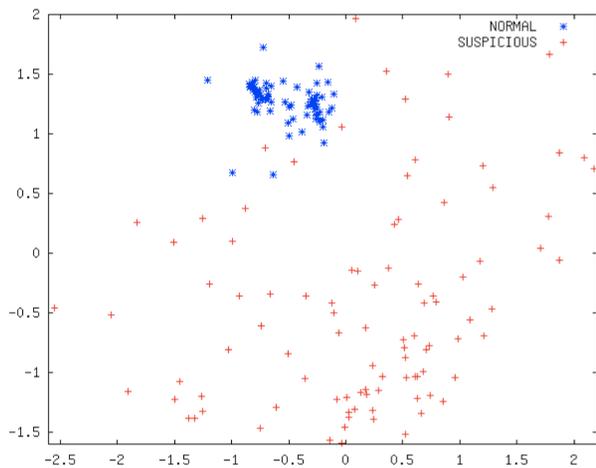
Currently, there is no standard ontology for AAA radiology reports. Consequently, the objective of this work is to develop an algorithm for learning the features of the language used in AAA reports.

## 2. DATA

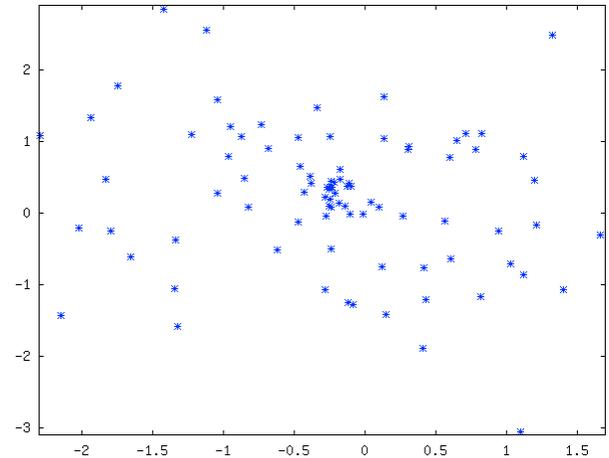
The data used in this research consists of 20 AAA patient records. There are a total of 111 reports consisting of 87 radiology reports and 24 operative notes, which are approximately 4 radiology reports and 1 operative note per patient. Patients undergoing AAA repair will have a pre-operative CT scan to visualize and describe the patient’s AAA. The patient then undergoes surgery to repair the AAA. Finally, the patient has several follow-up CT scans to check the progress of the repair. If the repair does not work, then additional surgery is performed along with additional follow-up scans.

### 2.1 Characteristics of the Data

As described in [6][7], mammography radiology reports are labeled as being either *normal* or *suspicious*. With the use of the BI-RADS system, these reports were further categorized according to the patient’s condition [1]. As a result, the language of the reports has unique characteristics according to their classification of either normal or suspicious. Normal reports tended to be shorter in length with less variability in the language used. Suspicious reports tended to be longer with wider variability in the language because they describe details of abnormalities in the patient. Figure 1 shows a multi-dimensional scaling of mammography reports using cosine distance and a TF-IDF vector representation [12]. As can be seen, the normal reports tend to cluster while the suspicious reports are widely scattered with little clustering.



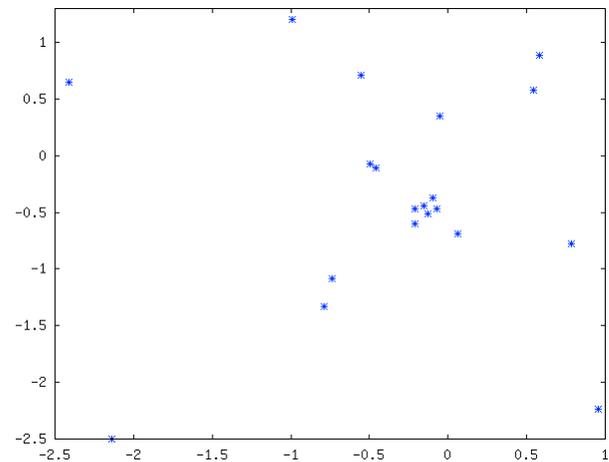
**Figure 1. Multi-dimensional scaling of mammogram radiology reports**



**Figure 2. Multi-dimensional scaling of AAA radiology reports**

In contrast, the AAA radiology reports do not have labels of *normal* or *suspicious*. The reports simply contain the observations of the radiologist about the patient. Figure 2 shows a multi-dimensional scaling of the AAA reports. As can be seen, the reports are widely scattered with a few reports clustering together.

The reports for each patient were also combined into a single patient record. There are, on average, 4 reports per patient record. Figure 3 shows a multi-dimensional scaling of the 20 patient records. As can be seen, the patients are widely scattered with a few patients clustering together.



**Figure 3. Multi-dimensional scaling of AAA patient records**

As will be discussed in a later section, the results of the multi-dimensional scaling are due to the characteristics of the language used to describe the AAA patients. While the AAA reports do not have a particular labeling scheme like mammography reports, they do have similar characteristics in that abnormalities in patients will result in reports that are both longer and contain wider variation in the language used. Patients with no abnormalities have shorter reports with less variability in the language.

The main problem of trying to identifying and analyzing reports describing abnormalities without a predefined labeling scheme or ontology lies in the language that is used in reports. Much like the

mammography reports discussed in [6], AAA reports describing abnormalities tend to have a richer vocabulary than normal reports (i.e., without abnormalities in the patient). In addition, normal reports tend to have a higher number of “negation” phrases. These are phrases that begin with the word “no” such as in the phrase “no evidence of endograft leak.” Consider the phrases shown in Table 1 and Table 2. These are the negation phrases that generally occur in normal reports and the ones shown here are samples of the variations that have been found.

**Table 1. Example phrases using "no" and "leak"**

no evidence of endograft leak
no obvious leak
no evidence of graft leak
no endovascular leak
no evidence of leak
no endovascular leak on delayed imaging
patent with no evidence of graft leak
no evidence to suggest endovascular leak

**Table 2. Example phrases using "no" and "fluid"**

no pelvic fluid
no free pelvic fluid
no free fluid in the pelvis
no significant pelvic free fluid
no free fluid
no pelvic free fluid
no free air free fluid

In contrast to the mammography reports discussed in [6], the abnormal AAA reports do not have a consistent set of terms or phrases to describe abnormalities. This is due to the higher variety of abnormalities that can be seen in AAA scans as opposed to mammography scans. This characteristics causes the analysis of AAA reports to be more difficult than mammography reports.

Considering the language variations shown previously, the task of retrieving those reports that represent abnormalities is daunting. The variations of terms and syntax create a combinatorial explosion while, semantically, these combinations tend to mean the same thing. Consequently, the traditional vector space model based on individual terms does not adequately capture the semantics used in medical documents. The use of phrases does adequately solve the issue of capturing the semantics, as there are numerous phrase variations to describe the same conditions of patients.

To address this challenge of capturing the semantics of the radiology reports, this work relies on skip grams to represent common phrase patterns. S-grams are word pairs in their respective sentence order that allow for arbitrary gaps between the words [2][3][8]. The s-grams for Table 1 are the words “no” and “leak.” This s-gram uniquely identifies a particular semantic in the language of AAA reports and enables the identification of all possible variations of such phrases. Higher-level patterns may then be formed from these s-grams to help create ontology.

Consequently, the goal of this work is to extend and apply the Maximum Variation Sampling Genetic Algorithm (MVS-GA) as described in [6][7] to the AAA radiology reports, and to assess its

performance in identifying abnormalities in patients and characterize the language in this sub-domain of radiology reports.

### 3. APPROACH

As discussed in the previous section and in [6], radiology reports exhibit two characteristics. First, reports describing abnormalities tend to be longer and have a wider variation in the language that is used. Consequently, these reports tend not to cluster with other reports. The second characteristic is that reports where there are no abnormalities use more negation phrases than abnormal reports. It is these two characteristics that we seek to exploit in this approach.

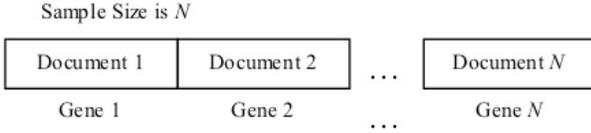
#### 3.1 Learning from Maximum Variation Sampling

The genetic algorithm used in this work is an extension of the one described in [6][7]. The genetic algorithm is used to implement maximum variation sampling, which helps identify critical and distinct features of the data set.

Maximum variation sampling is a nonprobability-based sampling. This form of sampling is based on purposeful selection, rather than random selection. The advantage of this form of sampling is that it allows a doctor or radiologist to look at data that may not otherwise be visible via the random selection process. Since abnormal mammography reports are not as common as normal ones, random sampling would make it difficult to find them. Within nonprobability-based sampling, there are several categories of sampling [5], one of which is maximum variation sampling (MVS) [5]. This particular sampling method seeks to identify a particular sample of data that will represent the diverse data points in a data set. In this case, the diverse data points will represent abnormal mammograms. According to [5], “This strategy for purposeful sampling aims at capturing and describing the central themes or principle outcomes that cut across a great deal of [data] variation.” The MVS is naturally implemented as a genetic algorithm (MVS-GA).

Before applying a GA to the analysis of mammography reports, the reports must be prepared using standard information retrieval techniques. First, reports are processed by removing stop words and applying the Porter stemming algorithm [4][9][10]. Once this has been done, the articles are then transformed into a vector-space model (VSM) [11][12]. In a VSM, a frequency vector of word and phrase occurrences within each report can represent each report. Once vector-space models have been created, the GA can then be applied.

Two of the most critical components of implementing a GA are the encoding of the problem domain into the GA population and the fitness function to be used for evaluating individuals in the population. To encode the data for this particular problem domain, each individual in the population represents one sample of size  $N$ . Each individual consists of  $N$  genes where each gene represents one radiology report (each report is given a unique numeric identifier) in the sample. For example, if the sample size were 10, each individual would represent one possible sample and consist of 10 genes that represent 10 different reports. This representation is shown in the following figure.



**Figure 4. Genetic representation of each individual**

The fitness function evaluates each individual according to some predefined set of constraints or goals. In this particular application, the goal for the fitness function was to achieve a sample that represents the maximum variation of the data set without applying clustering techniques or without prior knowledge of the population categories. To measure the variation (or diversity) of our samples, the summation of the similarity between the vector-space models of each document (or gene) in the sample is calculated as shown in the following equation.

In an effort to effectively characterize the phrase patterns of the AAA radiology reports, it is necessary to examine reports that are longer in length, so that more language can be examined for patterns. In addition, abnormal reports tend to be longer in length than normal reports since the radiologist is describing the abnormalities in more detail. To drive the GA towards longer reports and avoid duplicate reports, the fitness function of the MVS-GA incorporates penalty functions as shown in equations 1 – 4.

$$Fit(i) = \sum_{j=0}^N \sum_{k=j+1}^N \alpha_j + \beta_k + \chi_{jk} + Sim(Gene(i,j), Gene(i,k))$$

**Equation 1. Revised MVS-GA Fitness Function**

$$\alpha_j = e^{-\left(\frac{\|j\|}{100}\right)}$$

**Equation 2. Penalty factor for document  $j$**

$$\beta_k = e^{-\left(\frac{\|k\|}{100}\right)}$$

**Equation 3. Penalty factor for document  $k$**

$$\chi_{jk} = \begin{cases} 1,000 & j = k \\ 0 & j \neq k \end{cases}$$

**Equation 4. Penalty factor for documents  $j$  and  $k$**

In Equation 1, the Sim function calculates the similarity between the vector space models of gene  $j$  and  $k$  of the individual  $i$ . This similarity value ranges between 0 and 1 with 1 indicating that the two reports are identical and 0 indicating that they are completely different in terms of the words used in that report. Therefore, in order to find a sample with the maximum variation, Equation 1 must be minimized (i.e., lower fitness values are better). In this fitness function, there will be  $(N^2 - N) / 2$  comparisons for each sample to be evaluated.

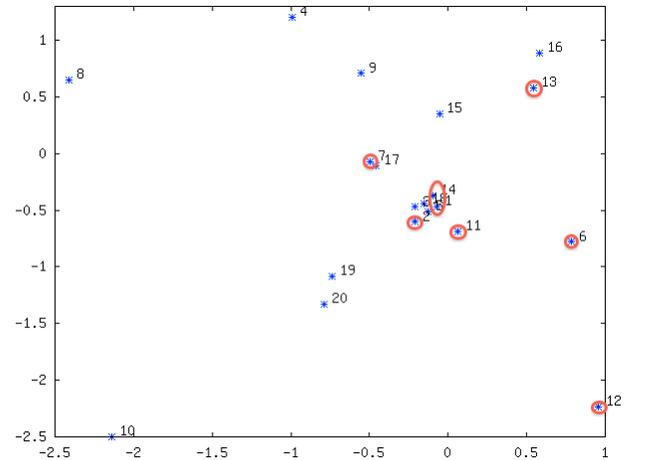
The penalty functions are incorporated into the fitness function in order to penalize individuals in the MVS-GA based on the length of the documents they represent. The penalty functions also return values that are between 0 and 1, inclusive. As a result of the penalty functions, the reports that do not discuss any complications or abnormalities tend to be shorter and receive the higher penalty values, while lengthy reports describing

complications will receive the lower penalty values. In addition, this work extends that of [7] by incorporating an additional penalty function that penalized for duplicate documents in the sample. The penalty value was arbitrarily chosen, but is sufficiently high enough to be a severe penalty.

After the MVS-GA is executed, the end result is a best sample of AAA reports that are as diverse from each other as possible. Once this sample is achieved, then phrases are extracted from each document in the sample. For each phrase in the document, s-grams are extracted. Next, the s-grams are counted across the sample of documents. S-grams that are common across the sample will have higher frequency counts while s-grams with a frequency of 1 uniquely identify a particular document in the sample.

## 4. RESULTS & CONCLUSIONS

Figure 5 shows the multi-dimensional scaling view of the patients that were selected by the MVS-GA. As expected, some of the patients are widely scattered, and the maximum variation sampling is designed to find widely scattered data points. However, some of the patients are clustered near each other. The reason for this is that they still have several distinct features in common, even though the overall content of these reports are very different.



**Figure 5. Patients selected by the MVS-GA**

The s-grams discovered by the MVS-GA from the data set are shown in Table 3 and Table 4. Table 3 shows the top negation s-grams from the best solution obtained. These negation s-grams are commonly used in the normal reports, but may occur in abnormal reports. They represent specific abnormalities that the radiologist is looking to find in the patient. For AAA patients, the radiologist is specifically looking to see if the repair of the aneurysm is not leaking and that there is no fluid collecting, or any other form of complication from the graft that inserted to repair the aneurysm. In addition to observing the AAA repair graft, the radiologist may also observe other abnormalities or the condition of other organs in the abdomen of the patient. This is the reason for the “no & mass” s-gram.

**Table 3. Top five negation s-grams from best solution obtained by MVS-GA**

S-gram	Example Phrase
no & leak	no evidence of leak
no & fluid	no pelvic free fluid
no & suggest	no finding to suggest
no & mass	no pancreatic mass
no & graft	no evidence of endovascular graft complication

The s-grams shown in Table 4 are the most common s-grams shared by the documents in the best solution obtained. As discussed previous, the MVS-GA searches to find documents that are as different from each other as possible. Consequently, any commonality between the documents in the sample is a result of the nature of the AAA procedure and patient condition. The first four s-grams in this table are a reflection of that. These are s-grams that most of the AAA patients are very likely to have in their reports, as they are standard descriptions for the procedures of the patients.

**Table 4. Top five common s-grams across the best solution obtained by the MVS-GA**

S-gram	Example Phrase
left & artery	left renal artery
pelvis & iv	pelvis after the administration of 142cc of omnipaque-300 iv
ct & abdomen	ct imaging of abdomen
hour & prep	3 hour oral prep
evidence & diverticulitis	sigmoid diverticulosis without evidence of diverticulitis

The last s-gram of Table 4 is particularly interesting. This s-gram does not refer to the AAA procedure, but is specific feature of the patients and refers to a condition known as sigmoid diverticulosis. Sigmoid diverticulosis is a condition where small pockets form along the wall of the colon where there are natural weak points. Currently, this condition has no connection to the AAA or any complications associated to the repair of AAA. However, in reviewing the results of the MVS-GA as shown in Table 5, it was found that patients with graft leaks have this condition. This condition also occurs in patients that do not have leaks or complications, but it is not as prevalent as those patients that do. Additional data and analysis will be needed to further investigate if this characteristic of the patients can be used in any way as a distinguishing feature of those patients that are likely to have complications after AAA repair.

Table 5 shows the general characteristics of the patients selected by the MVS-GA. Of these 8 patients, 3 of them had graft leaks, while the remaining patients have unusual characteristics that were noted by the radiologist.

One of the most significant aspects of these results is that the MVS-GA did not require any specialized ontology or dictionary or feedback from a subject matter expert. In addition, no search keywords, clustering, or categorization was needed. This allows medical researchers to search and identify interesting patient records without the need for explicitly defining what “interesting” patient records are.

**Table 5. Patient characteristics from the best solution obtained by the MVS-GA**

Patient	Abnormal Characteristics
2	Leak of the graft that heals on its own, sigmoid diverticulosis
5	Leak of the graft, sigmoid diverticulosis
6	Leak of the graft, sigmoid diverticulosis
7	No leak, possible pancreatitis, abnormal enhancement of the common bile duct
11	No leak, left renal artery stenosis
12	No leak, renal cysts, sigmoid diverticulosis, hepatic steatosis
13	No leak, suspected ulcer perforation of the stomach
18	No leak, ventral wall hernia, sigmoid diverticulosis

## 5. FUTURE WORK

While the work described here focuses primarily on the learning aspect of mining AAA patient reports, there are many avenues for future research. One area of further research will be a temporal analysis of the s-grams across the patient records. For medical treatment of AAA patients, it is necessary to know how the patient’s condition changes over time. The next area of future research is to investigate if characteristics such as sigmoid diverticulosis can be used as indicators of AAA repair complications. This is particularly important, as there is currently no indicator to forecast the success or failure of a patient’s recovery from AAA repair.

## 6. ACKNOWLEDGMENTS

Our thanks to Oscar H. Grandas, M.D. and the Department of Surgery, University of Tennessee Graduate School of Medicine and University of Tennessee Medical Center for providing the large dataset of images, radiology reports, and operating room notes on Abdominal Aortic Aneurysms, from which the test subset was chosen.

Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy. LDRD #5327.

This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

## 7. REFERENCES

- [1] American College of Radiology (ACR). ACR BI-RADS® - Mammography. 4<sup>th</sup> Edition. In: ACR Breast Imaging Reporting and Data System, Breast Imaging Atlas. Reston, VA. American College of Radiology; 2003.
- [2] Abdalla, R. M., and Teufel, S. 2006. A bootstrapping approach to unsupervised detection of cue phrase variants. In

- Proceedings of the 21<sup>st</sup> International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (Sydney, Australia). COLING 2006. ACM Press, New York, NY, 2061-2064.
- [3] Cheng, W., Greaves, C. and Warren, M. 2006. From n-gram to skipgram to conogram. *International Journal of Corpus Linguistics* 11/4: 411–33.
- [4] Fox, C. 1992. "Lexical analysis and stoplists." In *Information Retrieval: Data Structures and Algorithms* (ed. W.B. Frakes and R. Baeza-Yates), Englewood Cliffs, NJ: Prentice Hall.
- [5] Patton, M.Q. 1990. *Qualitative Evaluation and Research Methods*, Second Edition. Newbury Park, CA: Sage Publications, Inc.
- [6] Patton, R.M., Beckerman, B.G., and Potok, T.E., 2008. "Analysis of mammography reports using maximum variation sampling." *Proceedings of the 4th GECCO Workshop on Medical Applications of Genetic and Evolutionary Computation (MedGEC)*, Atlanta, USA, July 2008. ACM Press, New York, NY, 2061-2064.
- [7] Patton, R.M., Potok, T.E., Beckerman, B.G., and Treadwell, J.N., 2009. "A Genetic Algorithm for Learning Significant Phrase Patterns in Radiology Reports." *Proceedings of the 5th GECCO Workshop on Medical Applications of Genetic and Evolutionary Computation (MedGEC)*, Montreal, Canada, July 2009. ACM Press, New York, NY, 2061-2064.
- [8] Pirkola, A, Keskustalo, H., Leppänen, E., Käsälä, A. and Järvelin, K. 2002. "Targeted s-gram matching: a novel n-gram matching technique for cross- and monolingual word form variants." *Information Research*, 7(2) [Available at <http://InformationR.net/ir/7-2/paper126.html>]
- [9] Porter, M. 1980. "An algorithm for suffix stripping." *Program* vol. 14, pp. 130-137.
- [10] Porter Stemming Algorithm. Current Feb. 5, 2009. <http://www.tartarus.org/~martin/PorterStemmer/>
- [11] Raghavan, V.V., and Wong, S.K.M. 1986. "A critical analysis of vector space model for information retrieval." *Journal of the American Society for Information Science*, Vol.37 (5), p. 279-87.
- [12] Salton, G. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill.