Automatic Indexing of Handwritten Medical Forms for Search Engines

Robert Milewski, Venu Govindaraju

Automatic Indexing of Handwritten Medical Forms for Search Engines

Robert Milewski
University at Buffalo / CEDAR
520 Lee Entrance
UB Commons Suite 202
Amherst NY 14228
milewski@cedar.buffalo.edu

Venu Govindaraju
University at Buffalo / CEDAR
520 Lee Entrance
UB Commons Suite 202
Amherst NY 14228
govind@cedar.buffalo.edu

Abstract

A new paradigm, which models the relationships between handwriting and topic categories (denoted as 'concepts'), in the context of medical forms, is presented. The ultimate goals are (i) the recognition of medical handwriting, and (ii) the use of such information for a medical form search engine. Medical forms have diverse, complex and large lexicons consisting of English, Medical and Pharmacology corpus. This technique shows that a handwriting recognition engine, with just a few recognized characters, can be used to represent a medical concept. This allows (i) a reduced lexicon to be constructed, thereby improving the performance of handwriting recognition engines [6][21], and (ii) unseen PCR forms to be tagged with a concept and later searched. Both practical and theoretical numbers are reported. This research builds the notion of a 'computational semantic lexicon' which was vaguely introduced in our IWFHR 2002 paper [15] and incorporates other research in the area of call-routing [2][3].

Keywords: Lexicon, Medical, Handwriting, Search.

1. Introduction

The New York State Pre-Hospital Care Report (PCR)[20], has been obtained under an agreement with the Western Regional Emergency Medical Services (WREMS) division of the New York State (NYS) Department of Health. The recognition, lexicon reduction, and information retrieval algorithms and experiments, presented in this research, use these PCR documents (see Figure 1) [13][14][15][20].

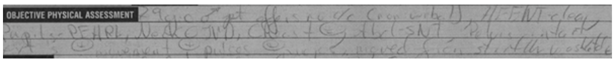
OBJECTIVE PHYSICAL ASSESSMENT

Figure 1 Medical Form Handwriting Example [20]

Handwriting recognition is used to tag medical forms, with a concept, for eventual searching and retrieval. The nature of the medical forms, involves large lexicons containing Medical, Pharmacology and English corpus. While current state of the art recognizers report recognition performance between ~58-78 percent, on comparable lexicon sizes [10][22], our execution shows ~25% raw match recognition performance. This illustrates the extremely complicated nature of medical handwriting. We have developed a method of automatically determining the concept of an unseen form using machine learning and computational linguistics techniques. We show how to use this paradigm to improve the word recognition, for a lexicon size of ~5000, by ~8% raw match rate. The basis for reducing the lexicon to improve recognition can be found here [6][21].

2. Lexicon Concept Hypothesis

This research proposes the following hypothesis which is verified experimentally: *A sequence of confidently recognized characters, extracted from an image of handwritten medical text, can be used to represent a concept. A lexicon can be reduced by keeping only those words belonging to those concept(s).*

2.1.1. Anatomical Concepts

Since the concept cannot be easily determined, a human is necessary to resolve the concept(s) for each training form. This work is dependent on the topology of words and concepts in the emergency medical domain. This anatomical topology, used as the PCR concepts, correspond to the patient ailment location(s). A PCR can be tagged with multiple concepts, but none had more then five tags.

11 Body Systems
Circulatory/Cardiac System
Digestive System
Endocrine System
Excretory System
Immune System
Integumentary System
Musculoskeletal System
Nervous System
Reproductive System
Respiratory System
Senses

6 Body Range Locations
Abdomen
Back/Thoracic/Lumbar
Chest
Head
Neck/Cervical
Pelvic/Sacrum/Coccyx

2 General
Full Body
Transported Patient

4 Extremities/Joint Locations
Arms/Shoulders/Elbows
Feet/Ankles/Toes
Hands/Wrists/Fingers
Legs/Knees

2.1.2. Classification

Unfortunately, the classification of patients is not an exact science; hence healthcare professionals are provided with guidelines which cannot replace experience. This ambiguous nature makes the construction of a hierarchical chart, representing all patient scenarios with respective prioritized anatomical regions, a difficult task which exceeds the scope of this research.

Example 1: A patient treated for an emergency pregnancy would be considered the *Reproductive System* concept.

Example 2: A conscious and breathing patient treated for gun shot wounds to the abdominal region would be considered *Circulatory/Cardiovascular System*, due to potential loss of blood, as well as other concepts such as Abdominal, Back, and Pelvic concepts.

3. Proposed Algorithm

The algorithm proposed here is derived from the Lucent Technologies research [2][3] involving the call-routing problem. Their strategy took voice recognition information as an input and produced the call destination as an output. In certain cases we deviate from their research to compensate for the differences in the research problems. However, this research shows that the lexicon reduction problem can be reduced to the call-routing problem.

In the training phase, a mechanism for relating uni-grams and bi-grams (hereon: uni/bi-grams) and concepts from a PCR training deck are constructed. The testing phase then evaluates the algorithm's ability to determine the concept, from an unseen form, by using a Lexicon Driven Word Recognizer (LDWR) [10] to extract the top-choice uni/bi-gram characters from all words. A maximum of two characters per word are trusted since the LDWR [10] will successfully extract a bi-gram with spatial encoding information 40% of the time. Trusting more then two characters by this LDWR [10] results in an excess of completely incorrect bi-grams.

3.1. Training

3.1.1. Filtering

Stopwords are omitted from the lexicon [16][7]. An additional list of ~50 words (e.g. 'male' 'female' etc...), found in most PCR's, which have no bearing on the concept, are omitted from the cohesion analysis in the next step, but will exist in the final lexicon. It is also common to apply other filtering methods on data to reduce the likelihood of morphological mis-matches [7]. However, such strategies as a stemming algorithm [7] cannot be applied before any recognition processes due to an additional layer of ambiguity; the text to retrieve is unknown. Consider a handwritten word image representing "rhythms" that needs to be recognized. The alteration of "rhythms" to "rhythm" in the lexicon, will effect recognition performance. However, at the end of classification, these words are considered equivalent. Therefore, word stemming is applied after the LDWR [10] has determined the ASCII word translation. This problem does not present in documents in which the text is actually known.

3.1.2. Phrase Construction

Diverging from the Lucent Technologies paper [2][3], word phrases and placeholders determination for separating a phrase is not used. The notion for defining a phrase as a sequence of adjacent non-stopwords can be found here [5]. Although an empirical study by Fagan indicated that important phrases may wrap around stopwords [5], the inclusion of stopwords degraded performance. In addition, since longer sequences of words, and longer sentences, were shown to be more successful then shorter contingent words [5], phrases are computed within the text area of 1 PCR region using a NLP cohesion technique used by Fagan [5][7].

A passage P is the set of all words w for a PCR form under a concept C treated as a single string. For each C, every pair of passages, denoted P_1 and P_2, are compared. Here we denote w_x as a word located at position x within a passage P. If $w_a \in P_1$ and $w_{a'} \in P_2$ and $w_b \in P_1$ and $w_{b'} \in P_2$ where $b > a$ and $b' > a'$, then a *potential phrase* consisting of exactly two words has been constructed. Once all potential phrases, under each C, have been determined, the cohesion of these phrases are computed. If the cohesion is above a threshold, then the potential phrase is considered a phrase that contributes to the representation for that concept C. Otherwise the potential phrase is thrown out. Therefore a concept C is represented by a sequence of phrases with high cohesion using only those PCR passages previously classified under C.

$$cohesion(w_a, w_b) = z \bullet \frac{f(w_a, w_b)}{\sqrt{f(w_a) * f(w_b))}} \qquad (1)$$

The cohesion between two words w_a and w_b is computed by the frequency that w_a and w_b occur together, denoted $f(w_a, w_b)$, divided by the square root of the independent word frequency of w_a, denoted $f(w_a)$, times the independent word frequency of w_b, denoted $f(w_b)$, which is then all multiplied by a constant z. In this research, z = 2 and the top 40 cohesive phrases are retained per concept (see equation (1)).

Phrase construction is a critical component to the development of the system both intuitively and computationally: (i) only those phrases which represent a concept should be used to model the concept, and (ii) a significant quantity of terms are discarded reducing time and space complexity. Consider the following as an example construction of a phrase under a concept:

Consider the two unfiltered text sentences S_1 and S_2 under the concept *Legs*:

S_1: "right femur fracture"
S_2: "broken right tibia and femur"

The candidate phrases CP_1 and CP_2 after the filtering step are reduced to:

CP_1: "right femur", "right fracture", "femur fracture"
CP_2: "broken right" ... "right femur" ...

A potential phrase from CP_1 and CP_2 will be computed as "right femur" since $w_{\mathbf{a}}$ and $w_{\mathbf{a'}}$ = "right", $w_{\mathbf{b}}$ and $w_{\mathbf{b'}}$ = "femur", and the condition b > a and b' > a' have been met. If the cohesion for "right femur" is above the threshold, across all PCR forms under the *legs* concept, then this phrase, representing the concept *legs*, is retained.

3.1.3. Term Extraction

All possible uni/bi-gram terms are then synthetically extracted from each cohesive phrase under each concept along with spatial information. For example, BLOOD can be encoded to the unigram 0B4 (zero characters before 'B' and four characters after 'B') and the bi-gram 0B3D0 (zero characters before 'B', three characters between 'B' and 'D' and zero characters following 'D'). All possible synthetic spatial encodings are generated for each phrase and chained together (a '$' is used to denote a chained phrase). For example, CHEST PAIN encodes to: 0C4$0P0A2 ... 0C4$1A2 ... 0C0H3$0P1I1 ... 0C0H3$0P2N0, etc... Therefore, each concept now has a list of encoded phrases, consisting of spatially encoded uni/bi-grams. These *terms* are the most primitive representative link to the concept used throughout the training process.

3.1.4. Term-Concept Matrix Construction

A matrix denoted A, of size |T| x |C|, is constructed such that the rows of the matrix represent the set of terms T, and the columns of the matrix represent the set of concepts C. The value at matrix coordinate (t,c), is the frequency that each term is associated with the concept.

Step 1 Compute the normalized matrix B from A using equation (2) [2]:

$$B_{t,c} = \frac{A_{t,c}}{\sqrt{\sum_{e=1}^{n} A_{t,e}^2}} \qquad (2)$$

Matrix A is the input matrix containing raw frequencies, Matrix B is the output matrix with normalized frequencies, and (t,c) is a (term, concept) coordinate within a matrix.

Step 2 Term Discrimination Ability

The popular TF*IDF (i.e. Term Frequency * Inverse Document Frequency) weighting approach is used to favor those terms which occur frequently with a small number of concepts, as opposed to their existence in all concepts. Two famous scientists in NLP, Luhn [12] and Salton [19] produced theories on the discrimination ability for terms and documents (i.e. 'concept' in this research). While Luhn [12] asserted that medium frequency terms would resolve a document the best, it precludes classification of more rare medical words. Salton's [19] theory, stating that terms with the most discriminate power are associated with fewer documents, allows rare-medium frequent word to resolve the document.

STEP 2A Compute the weighted matrix X from B using equation (3) [2][7]:

$$IDF(t) = \log_2 \frac{n}{c(t)} \qquad (3)$$

IDF computes the inverse-document-frequency on term t, and c(t) is the quantity of concepts containing term t.

Step 2B Weight the normalized matrix by IDF values using equation (4) [2][9][7]:

$$X_{t,c} = IDF(t) \cdot B_{t,c} \qquad (4)$$

Matrix B is the normalized matrix from Step 1, IDF is the computational step defined in Step 2, and Matrix X is a normalized and weighted matrix.

3.1.5. Reduced Singular Value Decomposition [4]

The normalized and weighted term-concept matrix can now be used as the knowledge base for later classification. A singular value decomposition variant, which incorporates a dimensionality reduction step, allows a large term-concept matrix to represent the PCR training set (5). This facilitates a concept query from an unknown PCR using the LDWR [10] determined terms later via [2][4].

$$X = U \cdot S \cdot V^T \qquad (5)$$

Matrix X is a matrix which is decomposed into 3 matrices: U is a (T × k) matrix representing term vectors, S is a (k × k) matrix, and V is a (k × C) matrix representing concept vectors.

The value k represents the quantity of dimensions to keep; hence the matrices are operating in k-dimensional space. If k equals the quantity of concepts to model, then the SVD is performed without the reduction step. Therefore, in order to reduce the dimensionality, the condition k < |C| is necessary. The theory behind the reduction is that the collapse of dimensional space can reduce noise [4].

3.2. Testing

Given an unknown PCR form, the task is to determine the concept of the form, and use the reduced lexicon associated with the determined concept to drive the LDWR [10].

Similarly with the call-routing paper [2][3], the query task is broken up into the following steps (with the precursor of *Binarization* to the *Term Extraction* step):

- Term Extraction
- Pseudo-Concept Generation
- Candidate Concept Selection

3.2.1. Binarization

PCR handwriting regions are extracted from a carbon copy down sampled grayscale document using the binarization and post-processing algorithm in [13] (compare Figures 1 and 2). Recognition performance is dependent on the clarity and solidarity of this handwriting information.

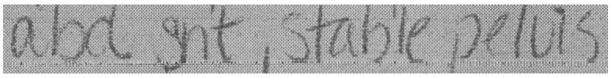

Figure 2 Carbon Copy Handwriting Example [13]

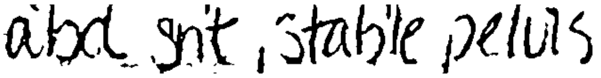

Figure 3 Binarized and Processed Text [13]

3.2.2. Term Extraction

Given a new PCR image, all image words are extracted from the form, and the LDWR [10] is used to produce a list of confident characters for each word. Only the most confident characters are used to encode the spatial uni/bi-grams consistent with the format during training. All combinations of uni/bi- phrases, in the PCR being evaluated, are constructed. Each word will have exactly one uni-gram and exactly one bi-gram; a phrase will consist of exactly two unknown words, and therefore be represented by precisely four uni/bi-phrases (BI-BI, BI-UNI, UNI-BI and UNI-UNI).

3.2.3. Pseudo-Concept Generation

An m x 1 query vector Q is produced, which is populated with the term frequencies for the generated sequences from the *Term-Extraction* step. If a term was not encountered in the training set, then it is thrown out. Spatial bi-grams are generated and found as trained terms 37% of the time, and similarly, spatial uni-grams 57% of the time. The experiments illustrate this to be a sufficient quantity of terms (see Section 6). A scaled vector representation of Q is then produced by multiplying Q^T and U.

3.2.4. Candidate Concept Selection

The task is now to compare the pseudo-concept vector Q with each vector in $V_r \bullet S_r$ (from the training phase) using a scoring mechanism. Consistent with [2][3], the cosine score is used for matching the query. Both x and y are dimensional vectors used to compute the cosine in the following equation (6):

$$z = \cos(x,y) = \frac{x \cdot y^T}{\sqrt{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2}} \qquad (6)$$

Each cosine score is then mapped to a point on a sigmoid function, using the least square fitting method, thereby producing a more accurate confident score [2][3]. The least squares regression line equations used to satisfy the equation *f(x) = ax + b* are (7) (8) [11]:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad (7)$$

$$b = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i\right) \qquad (8)$$

The fitted sigmoid confidence is produced, using the cosine score and the regression line, via the equation (9):

$$confidence(a,b,z) = \frac{1}{\left(1 + e^{-(az+b)}\right)} \qquad (9)$$

The sigmoid confidence scores, one for each concept, are then used to rank the chosen concepts in order of preference. The rank is then thresholded, and all words under the selected concepts are used to construct a new lexicon, which is then submitted to the LDWR recognizer [10]. Given a PCR in question, and this newly reduced lexicon, the LDWR [10] iterates though all image representations of handwritten medical words producing an ASCII interpretation.

3.2.5. Result and Truth Comparison

Each word which is recognized (i.e. the result) is finally compared to the human classification of the word (i.e. the truth) to determine performance. However, a simple string comparison is insufficient due to spelling mistakes and root variations of word forms which are semantically identical. This occurs 20% of the time within the test deck words. Therefore, a Porter stemming [17][8][18] and a Levenshtein String Edit Distance [1] of 1 allowable penalty are performed on both the truth and result before they are compared. Levenshtein is only applied to a word that is believed to be ≥ 4 characters in length. For example, PAIN and PAINS are identical. However, this also results in an improper comparison in ~11% of the *corrections*. These are the words incorrectly classified as equivalent:

FIGHT vs EIGHT vs LIGHT
MEDICAL vs MEDICATION
1400 vs 2400
BAIL vs RAIL
BALL vs CALL
MOLE vs MOVE
CALF vs CALL
CARD vs CARE vs CART
COLD vs TOLD
NECK vs DECK
FALL vs CALL
FEET vs FEED
FOUND vs BOUND vs SOUND vs POUND

FINE vs FIRE
FOOD vs FOOT
LEFT vs LIFT
MOANING vs MORNING
MARK vs MARY
PUNCH vs LUNCH
REACH vs REACT
SCARE vs CARE
SEVER vs FEVER
STABLE vs TABLE

3.2.6. Medical Form Search Engine

Finally, a set of unknown medical forms can be automatically classified into their appropriate concept(s). After all forms have been tagged with a concept, an authorized user can then supply a series of keywords as input, which is then compared to the cohesive phrases of the known concept(s). The PCR's matching those concept(s) are then returned to the user.

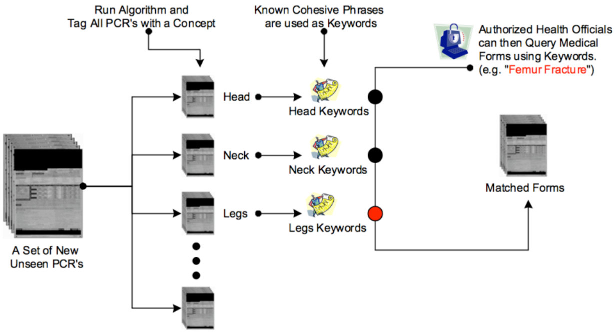
Run Algorithm and Tag All PCR's with a Concept
Known Cohesive Phrases are used as Keywords
Authorized Health Officials can then Query Medical Forms using Keywords. (e.g. "Femur Fracture")
Head
Head Keywords
Neck
Neck Keywords
Legs
Legs Keywords
Matched Forms
A Set of New Unseen PCR's

Figure 4 Concept Tagging and Query Engine

4. Results

In this section, the results of several experiments illustrate the effectiveness of this algorithm. Accept rate, error rate, and raw rates are reported for each experiment.

Table 1 Handwriting Recognition Performance

	CL	CLT	AL	ALT	RL	RLT
ACC	74.93%	75.32%	62.48%	66.44%	69.40%	70.18%
ERR	70.49%	67.98%	55.32%	43.45%	59.61%	57.20%
RAW	24.57%	26.86%	33.33%	44.44%	31.99%	34.34%
TLS	5,029	6,561	1,080	1,115	2,449	2,520
!L	0%	0%	24.62%	6.58%	14.57%	10.88%
!HL	-	-	25.85%	96.61%	43.30%	57.94%

Table 2 Environment

Training Deck PCR Size	619
Testing Deck PCR Size	40
Training Deck Words	5,029
Testing Deck Words	1,791
Training + Testing Deck Words	6,561
Modeled Concepts	23
Concepts used in Lexicon Reduction	5
Maximum Concepts per Form	5
Average Concepts per form	1.4
Cohesion Threshold Per Concept	Top 40
Apple OS X Memory Usage	400 MB
Apple OS X G4 1GHZ Runtime	90 mins

PERFORMANCE MEASURES

ACC (accept recognition rate)
This value reports how many words the LDWR [10] was confident in accepting (i.e. those words above an empirically decided threshold).

ERR (error recognition rate)
This value reports how many LDWR [10] accepted words (i.e. those words above an empirically decided threshold) were incorrect accepts.

RAW (raw recognition rate)
This value represents the performance ignorant of accept/reject values. This rate is the rate for which the top choice LDWR [10] word matched the truther word.

TLS (total lexicon size)
The size of the lexicon either complete or by reduction.

!L (truther word not present in the lexicon)
This value indicates the percentage of words, for a specific experiment, not in the lexicon as a result of incorrectly chosen concept(s) or the absence of that word in the training deck; depending on the experiment.

!HL (human could not completely decipher word)
This value indicates the percentage of those values in the !L set in which the human could not reasonably decipher all or some of the characters in the word.

EXPERIMENTS

CL (complete training lexicon)
The complete lexicon, which is the union of all words in the training set, is submitted to the LDWR [10].

CLT (complete training lexicon + test deck lexicon)
The complete lexicon, which is the union of all words in the training and test sets, are submitted to the LDWR [10].

AL (assumed training lexicon)
This is a reduced lexicon from the training deck such that the concepts are assumed to be determined by an Oracle. Only words from those assumed concepts are used in the lexicon construction.

ALT (assumed training lexicon + test deck lexicon)
Same as AL except that all words from the test set are inserted into the training deck concept lexicon. This shows the best case theoretical upper bound for the effectiveness of the reduced lexicon strategy.

RL (reduced lexicon)
The reduced lexicon from the training deck, which is the union of words from the top 5 ranked concepts is submitted to the LDWR [10]. *This is a practical measure of the performance of the system.*

RLT (reduced lexicon + test deck lexicon)
Same as RL except that all words from the test set are inserted into the training deck concept lexicon. This shows the effectiveness under the assumption that the concept lexicons are complete.

DISCUSSION

The theoretical RLT (i.e. comparing RLT to CLT) improves the RAW match rate by 7.48% and drops the error rate 10.78%, while removing 61.59% of the lexicon words.

The practical RL (i.e. comparing RL to CL) improves the RAW match rate by 7.42% and drops the error rate 10.88%, while removing 51.30% of the lexicon words.

The reason why the RLT and RL numbers are close is due to the different in the initial lexicon sizes: CLT/RLT starts with 6,561 words (i.e. training deck and testing deck lexicons) whereas the CL/RL starts with 5,029 words (i.e. training deck lexicon only). The RLT lexicon is more complete, but the lexicon will be larger. The RL lexicon is less complete, but the lexicon will be smaller. Both ways show a benefit: RLT benefits b/c the recognizer has a greater chance of the word being a possible selection; the RL benefits with the lexicon being smaller, therefore a word already in the lexicon has a greater chance of being selected. This dual benefit shows strength in the scalability of the paradigm.

The ALT shows the maximum theoretical upper bound for the paradigm: (i) the concepts are correctly determined 100%, and (ii) the lexicon is complete. The ALT (i.e. comparing ALT to CLT) improves the RAW match rate by 17.58% and drops the error rate 24.53%, while removing 83.01% of the lexicon words.

5. Future Work

In order to accomplish a completely operational health surveillance system, the following tasks remain:

- A form dropout and registration system.
- A word location and segmentation algorithm.
- The integration of a symbol recognizer.
- The inclusion of semantic stemming/spelling.
- Automated concept determination.
- Paradigm scalability using a nationally sampled training and test deck.

6. Conclusions

This paper defines a new paradigm for lexicon reduction in the complex situation of handwriting recognition of medical forms. The strategy is novel in its hybridization of linguistics, statistical modeling and handwriting recognition. A series of theoretical and practical recognition rates are provided as evidence. An improvement of ~7.5-17.6% raw match rate, ~11-25% reduction in error rate, and ~50-80% reduction in lexicon size have been shown in these practical and theoretical experiments. Basic intuition into human cognitive processes during the recognition of a word, in unknown and unclear contexts, can be seen.

7. References

[1] Black, P.E., ed. "Levenshtein Distance". Algorithms and Theory of Computation Handbook; CRC Press LLC, from Dictionary of Algorithms and Data Structures, NIST. 1999.

[2] Chu-Carroll, J., and Carpenter, B., Vector-Based Natural Language Call Routing. *Computational Linguistics*. Vol. 25, No. 3, pp. 361--388, 1999.

[3] Chu-Carroll, J., and Carpenter, B. Dialogue Management in Vector-Based Call Routing. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*. pp. 256--262, 1999.

[4] Deerwester, S., Dumais, S.T., Furnas, G.Q., Landauer, and, T.K., Harshman, R. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*. 41(6):391-407, 1990.

[5] Fagan, J. The Effectiveness of a Non-Syntactic Approach to Automatic Phrase Indexing for Document Retrieval. *Journal of the American Society for Information Science*, 40: 115-132. 1989.

[6] Govindaraju, V., Slavik, P., and Xue, H. Use of Lexicon Density in Evaluating Word Recognizers. *IEEE Trans PAMI*, Vol. 24, No.6, p.789-800. 2002.

[7] Hersh, W.R. Information Retrieval: A Health and Biomedical Perspective. 2nd Edition. *Springer-Verlag*, New York, Inc. USA. 2003.

[8] Jones, K.S. and Willet, P.. *Readings in Information Retrieval*, San Francisco: Morgan Kaufmann. 1997.

[9] Jones, K.S.. A Statistical Interpretation of Term Specificity and its Application in Retrieval. *Journal of Documentation*, 28(1):11-20 1972.

[10] Kim, G., and Govindaraju, V. A Lexicon Driven Approach to Handwritten Word Recognition for Real-Time Applications. *IEEE Trans. PAMI* 19(4): 366-379 1997.

[11] Larson, Hostetler, Edwards. Calculus with Analytic Geometry. Chapter 13 : Section 13.9. Fifth Edition. D.C. Heath and Company. USA. C1994.

[12] Luhn, H. A Statistical Approach to Mechanized Encoding and Searching of Literary Information. *IBM Journal of Research and Development*, 1: 309-317. 1957.

[13] Milewski, R. and Govindaraju, V. Extraction of Handwritten Text from Carbon Copy Medical Form Images. *International Workshop on Document Analysis Systems*. 2006.

[14] Milewski, R. and Govindaraju, V.. Handwriting Analysis of Pre-Hospital Care Reports. *Proceedings of the 17th Symposium on Computer-Based Medical Systems* (IEEE CBMS). 2004.

[15] Milewski, R. and Govindaraju, V. Medical Word Recognition using a Computational Semantic Lexicon. *Eighth International Workshop on Frontiers in Handwriting Recognition*. Canada. 2002.

[16] National Library of Medicine. PubMed Stop List.

[17] Porter, M.F.. An Algorithm for Suffix Stripping. *Program*, 14: 130-137. 1980.

[18] Rijsbergen, C.J. van, Robertson, S.E. and Porter, M.F.. New Models in Probabilistic Information Retrieval. London: British Library. 1980.

[19] Salton, G. Introduction to Modern Information Retrieval. New York. McGraw-Hill. 1983.

[20] Western Regional Emergency Medical Services. Bureau of Emergency Medical Services. New York State (NYS) Department of Health (DoH). Prehospital Care Report v4.

[21] Xue, H., and Govindaraju, V. On the Dependence of Handwritten Word Recognizers on Lexicons. *IEEE Trans. PAMI*, Vol. 24, No. 12, p. 1553-1564. 2002.

[22] Xue, H. and Govindaraju, V.. Stochastic Models Combining Discrete Symbols and Continuous Attributes – Application in Handwriting Recognition. *Proceedings of 5th IAPR International Workshop on Document Analysis Systems*. pp. 70-81. 2002.