Application of Some Retrieved Information Method on Internet

Vu Thanh Nguyen[1], Nguyen Quoc Thinh[1]

[1]University Of Information Technology, 34 Truong Dinh Str., 3 Dist., HoChiMinh City, VietNam

Abstract

This paper compares several methods of information extraction on the internet. Today, internet has become a treasure of knowledge. Every year, thousands of pieces of different information are posted on the internet. So, extracted information on the internet for many different purposes has become an important problem today. Users may extract information based on some available tools such as Lapis, Risk, Rapier, Wien, and Stalker… However, these tools have a disadvantage: we must update the training data when the website changes. So SVM and CRF associated with natural language processing are the best solutions to solve this problem. Information extraction from online Vietnamese news website with SVM and CRF brings experiment results that is very optimistic. Its results reach nearly 90% of the accuracy in websites and the processing time is less than one minute per site when the specified number of link levels is 1 within the base site.

Keywords: RI (Retrieved Information), CRFs (Condition Random Fields), SVM (Support Vector Machine), ECT (Embedded Catalog Tree).

I. Introduction about retrieved information on the internet

Nowadays, the internet is a huge library of the whole world, with a lot of information that includes all fields, all jobs in industry, agriculture, economy, finance…So this retrieved information has become very imperative. We need to solve the problem: How is realized and retrieved information extracted as exactly as possible from the internet?

II. Some methods for solving this problem

Hand-code wrapper [5]: the first wrapper was created is hand-code wrapper, this method uses the consistence of web pages to create its wrapper. For example, information from showtimes.hollywood.com can be retrieved easily by a simple structure: information like show-time, movie title… always stays in special tags (show-time always is bold with tag <B></B>).

The advantage of this method is simple and this wrapper is created easily. However hand-code wrapper is only used to present some consistent web pages and each wrapper is only used for one web page. Nowadays, with the continous changing of web pages, it is useless to use hand-code wrapper, that's also the reason why automatic wrapper construction was taken from.

Automatic Wrapper Construction (AWC) [5][6]: With this method, wrapper was created automatically based on some requests of users, such as information on what field, what career, …And each wrapper can be used for a group of web pages (with similar structure).

AWC-LR Wrapper Class [5][6]: LR (Left Right delimiters) is a set of delimit symbols on left and right side of field of needed information, input for this method is only the address of web pages which have marked the field of needed information, and after that it will try all of cases of delimit to create a suitable wrapper.

The advantage of this method is easy to use, we just add addresses of web pages, the rest tasks will be automatic and receive wrapper for web pages. However, this method can only operate perfectly with web pages, where information have consistent delimit symbol and at the same time information is only realized based on LR delimiters, so it is easy to meet an error.

For example:

<HTML><TITLE>Country Codes</TITLE>

<BODY><B>Congo</B> <I>242</I>

<B>Spain</B> <I>34</I>

<HR><B>END</B></BODY></HTML>

We determine (<B>, </B>, <I>, </I>) with LR Wrapper Class.

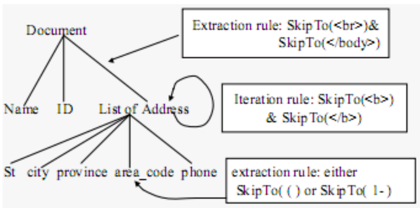
AWC-HLRT Wrapper Class [5][6]: H(Head delimiter), LR(Left Right delimiter), T(Tail delimiters) are delimit symbols at head, left, right and tail of information we want to retrieve. This is an improved model of LR Wrapper Class, so its advantage and disadvantage are the same. There is just a different thing, this method has Head and Tail delimit symbol at bonus, and therefore retrieved information can be marked more exactly.

Embedded Catalog Tree (ETC) [5]: Leaves are the items users want, local nodes (even root) represent a list of data sets, each item in any data set may be a leaf l or a list L (Embedded List) of k which lets we know the number of items in data sets.

Preparing with the two above methods, LR and HLRT Wrapper Class are more improved with more exact information retrieving. Proceeding from ideas, information in web pages was represented based on level and data sets. When user adds web pages, which field of want-to-retrieve information was marked, with this field and ECT method, a new rule set was born. Retrieving process will execute with path P form root to suitable leaf and retrieve every x which belongs to P from its father node. Rule set was presented by Landmarks, Wildcard, Function, Cascade Function, Selection Rules.

Landmarks: group of continuous tags (like works, numbers, HTML Tags, substring,…) for example: <b>,<p>…

Wildcard: class of tags (like number, Sign, HtmlTag, Allcaps,…)
Function: an expression with a landmark or a wildcard is variable. For example: SkipTo(:<b>)…
Cascade Function: like SkipTo(Allcaps), NextLandmark(number)…
Selection Rules: like SkipTo(<b>) or SkipTo(<i>)



Its advantage is that we can retrieve information exactly up to 90%. However, a lot of web pages always change content and way to show information, accordingly wrapper has been updated regularly. This also is mistake of almost wrappers nowadays.

III. Retrieved information on news web pages with method combined Condition Random Fields (CRFs) with Natural Language Processing (NLP)

CRFs is a model of undirected linear state (a trained limited state engine with conditions) and follows the first Markov property. CRFs was proved successfully for labels assign projects like separate words, assign labels for phrase, determine entities, assign labels for group of nouns, etc…
Call $\mathbf{o} = (\mathbf{o_1}, \mathbf{o_2}, \ldots, \mathbf{o_T})$ a sequence of observed data which will be assigned labels. Call S a state set, each state set is linked with a label $l \in L$. Let set $\mathbf{s} = (\mathbf{s_1}, \mathbf{s_2}, \ldots, \mathbf{s_T})$ be a certain sequence of state, CRFs determines condition probability of sequence of state follow this function:

$$p_\theta(\mathbf{s} \mid \mathbf{o}) = \frac{1}{Z(\mathbf{o})} \exp\left[\sum_{t=1}^{T} \sum_{k} \lambda_k f_k(s_{t-1}, s_t, \mathbf{o}, t)\right].$$

Call $$Z(\mathbf{o}) = \sum_{s'} \exp\left(\sum_{t=1}^{T} \sum_{k} \lambda_k f_k(s'_{t-1}, s'_t, \mathbf{o}, t)\right)$$

standardized factor for all of possible labeled sequences. f_k determines a specific function and λ_k is a linked variable that links with each f_k. The purpose of machine learning of CRFs is to measure these variables. Therefore, we have two kinds of specifics f_k : state specific (per-state) and transition specific(transition).

$$f_k^{(per-state)}(s_t, \mathbf{o}, t) = \delta(s_t, l) x_k(\mathbf{o}, t). \quad (2)$$

$$f_k^{(transition)}(s_{t-1}, s_t, t) = \delta(s_{t-1}, l)\delta(s_t, l). \quad (3)$$

Here δ is Kronecker-δ . Each per-state (2) combines label l of current state s_t and a contextual predicate – a binary function $x_k(\mathbf{o},t)$ determines the important contexts of observe $\mathbf{o}$ at time t. A transition specific (3) shows the dependent sequence combined with label l' of previous state s_{t-1} and label l of current state s_t. Usually, training CRFs is proceeded by maximum likelihood training data along with optimal technology such as L-BFGS. The argument (based on learned model) is finding out correlative label sequence of input observe sequence. With CRFs, usually the use of traditional motion scheming algorithm is Viterbi to argue with new data.

III.1. Illustrate problems

a. <u>Goal</u>

- Determine websites which contain news or not?
- Determine information fields which have news?
- Classify news?

b. <u>Solve the problem</u>

With a web site x and DOM set (document object model), nodes $x^1,\ldots,x^k$ in x. Set $\vec{y} = y^1,\ldots,y^k$ is labels that can be assigned to $x^1,\ldots,x^k$ with $y^1,\ldots,y^k$ are contents of nodes $x^1,\ldots,x^k$. Examine in turn specific nodes, which include contents, to examine labels and choose label which have content, news we need.

c. <u>Appeared problems</u>

- Determining websites which contain news means that when adding any websites, we have to find out whether this website contains news or not?
- If this website contains news, search in data fields where news is contained.
- By "had news" field, determine what kind of news is?

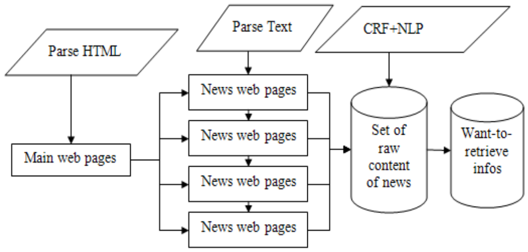
d. <u>Difficulties</u>

- Determining "had news" field bases on content of news, this work relate to problem Vietnamese document processing, a tough problem.
- Processing websites is performed through a lot of steps which are easy to appear potential error.

e. <u>Advantages</u>

- Vietnamese processing has just happened at determining word-class (PosTagging), this simplifies problems.
- Determining grammar of sentences has just stopped at step "Subject - Predicate" determining.
- Help of Vietnamese dictionaries of about 30.000 words.

III.2. Some basic algorithms on combined CRFs and NLP



H1. Map of CRF and NLP

Determining news web pages

Input: Address of web page needs to be tested and a list of sample web pages (training data set)

Output: Answer for Question : This web page is news web page or not?

Algorithm: web pages in sample web page set

Determine field of information

Input: want-to-retrieve web page

Output: link news web pages and its content

Algorithm:

- Parse text
- Use combine CRFs and NLP to find content of news (based on specifics of data field)
  - o Separating Segmental to words and count number of word
  - o Assigning labels word-class for words, count number of noun
  - o Checking conditions of content field which has news

III.3. Assigning label processing

First step: Determining word-class labels for suitable word which bases on syntax rules and contexts (noun, verb,..about 48 word-classes)

Second step: Starting to assign labels, find out all possible word-class labels for each remain word.

Third step: Deciding to assign label results, removing unclear.

Assigning label by probability method [1][2][3]

Ideas

Determining distributive probability in combined space between chain of words Sw and chain of word-class St

Removing unclear word-class problem for a chain of words that can be transferred to a chain of word-class selecting such that conditional probability $P(S_t|S_w)$ can combine these chains of word-class with maximum value chain of words.

Process

Follow Bayes formula:

$$P(S_t \, |S_w) = P(S_w|S_t).P(S_t)/P(S_w)$$

With each $St = t_1t_2 \ldots t_N$ and each $S_w = w_1w_2 \ldots w_N$:

$P(w_1w_2\ldots w_N \mid t_1t_2\ldots t_N) = P(w_1 \mid t_1t_2\ldots t_N)\, P(w_2 \mid w_1,t_1t_2\ldots t_N)\ldots P(w_N \mid w_1\ldots w_{N-1}, t_1t_2\ldots t_N)$
$P(t_1t_2\ldots t_N) = P(t1)P(t_2 \mid t_1)\, P(t_3 \mid t_1t_2) \ldots P(t_N \mid t_1\ldots t_{N-1})$

Each P(wi | w1... wi-1, t1t2...tN), suppose, appearing chance of a word, when knowing word-class label, is defined completely if we know that label. This means that $P(w_i|w_1\ldots w_{i-1},t_1t_2\ldots t_N) = P(w_i \mid ti)$. So, probability $P(w_1w_2\ldots w_N|t_1t_2\ldots t_N)$ only depends on basic probabilities $P(w_i|t_i)$:

$$P(w_1w_2\ldots w_N \mid t_1t_2\ldots t_N) = P(w_1 \mid t_1)P(w_2 \mid t_2) \ldots P(w_N \mid t_N)$$

With probabilities $P(t_i \,|t_1\ldots t_{i-1})$, suppose, chance to appear of a word-class is totally clear, if we know label of word-class in a gap which has a fixed size k. This means that $P(t_i| t_1\ldots t_{i-1}) = P(t_i|t_{i-k}\ldots t_{i-1})$. In general, assigned label tools often suppose k with 1(bigram) or 2 (trigram).

With a library of labeled documents, parameters of this model are determined easily by Viterbi algorithm.

Viterbi Algorithm

With a sequence of words $W_1, \ldots, W_T$, word-class $C_1, \ldots$, CN, probability Pr $(W_i \mid C_i)$ and probability Bigram Pr $(C_i| C_j)$, find a sequence of word-class $C_1,\ldots,C_T$ with the best suitable for a sequence of words $W_1,\ldots, W_T$.

Begin:

for i = 1 to K do
SeqScore(i,1) = Pr(C1 | <Start>)* Pr(W1 | Ci);
BACKPTR(i,1) = 0;
Repeat:

for t = 2 to T do
for i = 1 to K do
SeqScore (i,t) = Max (SeqScore (j,t -1)* Pr (Ci | Cj))* Pr (Wt | Ci), với j = 1,..K;
BACKPTR(i,t) = maximum value of j
Determining a sequence of word-class:

C(T) = i is Max of SeqScore(i,t);
for i = T-1 to 1 do
C(i) = BACKPTR(C(i+1),i+1);

IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 6, November 2010
ISSN (Online): 1694-0814
www.IJCSI.org

Assigning labels based on literary style [1][2][3]

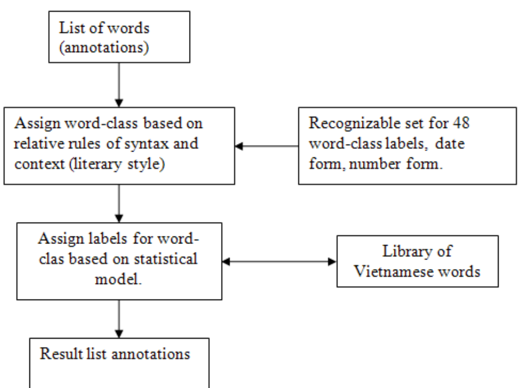
Following the way that the documents represent in a specific context to determine word-class for words, and even the determining makes sure that grammar rules of words in sentence are still right.

Represent through procedures: Using rules of proper nouns determining. According to determined proper nouns, continue using rules to determine 48 remain word-class labels.

Assigning labels based on the combine of literary style and probability [1][2][3]

Assigning label engine is a mixed system which is based on literary style and assigning label engine trigram.

Processing



H2. Processing of word-class assign labels

Assigning label engine working with input data is a list of annotations, each annotation is linked with a word in document. Assigning label engine can assign a sequence which consists of 4 word-classes with probability information for each word in list, or only assigning the last result – labels have the highest chance to appear. At last, we got annotations with constructed like:

Algorithm

Read all words in document;

- Assigning word-class labels for words without confusion;
- Using rules of proper nouns determining;
- According to determined proper nouns, continue using rules to determine 48 remain word-class labels;

o Write to buffer;

o While(buffer not empty) do

Read 3 words from buffer;

- for each word from these 3 words do
  - if it's in dictionary then assign it with all labels (tag) in dictionary;
  - else assign it all possible labels (tag);
  - j = 0;
  - while(j < number of labels) do
    - o Counting P_w = P(tag|token) is probability from token with tag label;
    - o Calculating P_c = P(tag|t_1,t_2), is probability of tags which appear behind the labels t_1, t_2, are suitable labels of 2 words, which is in front of token word;
    - o Calculating $P_{w,c} = P_w * P_c$, combines 2 above probabilities(P_w and P_c);
    - o j = j +1;
  - end while;
- end for;

o end while;

III.4. Specific of news content field

- News usually stay on tag [P][/P][a][/a]
- Length of content of news > 50 and number word (after separated segmental) > 30
- Number of verbs (after assign labels PosTagger) > 2 (at least 2 sentences)
- Number of nouns (after assign labels PosTagger) > 10

IV. Retrieved information on new web pages with Support Vector Machine (SVM) [9][10][11]

Support Vector Machine (SVM) is a classified method in machine learning, it was proposed by Vladimir and co-workers from 1970s. SVM's origine was from statistical theory, based on Structural Risk Minimisation, which worked with the idea: keeping Test Error is minimum. When it's processing, SVM will move the sample set from presenting space R^n to space R^d with the higher of number of dimensions. In space R^d, find a optimal super plane to separate this sample set based on its classification, it means, find out distributive field of each class in space R^n, and determine classification of subclass of sample. Although, this method has improved from 1970s, scientists only truly care about SVM when the first paper was placed since 1995.

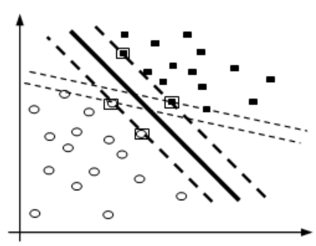
Application of SVM has been encounted a lot in "recognize hand-writing words" (Cortes and Vapnik, 1995; Scho¨lkopf, Burges and Vapnik, 1995; Scho¨lkopf, Burges and Vapnik, 1996; Burges and Scho¨lkopf, 1997), recognize objects (Blanz et al., 1996), recognize face in picture (Osuna, Freund and Girosi, 1997), document classification (Joachims, 1997).

Therefore, retrieved information from web with SVM is also executed by using document classification.

Main idea of algorithm Check any document a_i in document set a which belongs to classified A or not (information classification)? If $a_i \in A$ then a_i is assigned label 1, else d is assigned label -1. This processing method repeats until all of documents in document set are classified and want-to-retrieve information is document assigned with the label 1. This method classifies document to get right field of information which user want, the rest is only like getting a gift on table.

Algorithm [10][11] Suppose we have a set of specifics $T=\{t_1, t_2, \ldots, t_n\}$, each document a_i is represented by a data vector $x_i=(w_{i1}, w_{i2}, \ldots, w_{in})$, $w_{ij} \in R$ is variable of word t_j in document a_i. Hence, position of each data vector x_i correlate with position of a point in space R^n. Processing of classification will be executed on data vector x_i , it's not from document d_i . Therefore, we will use identically vocabularies, document, data vector, data point.

Training data of SVM is labeled documents $Tr=\{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$(this is set of news document which was collected from news web pages), let x_i be data vector, represents document d_i ($x_i \in R^n$), $y_i \in \{+1, -1\}$, (x_i, y_i). This means that vector x_i (or document d_i) is assigned label y_i . If we consider each document d_i as data point in space R^n , SVM will find the best geometric surface (super plane) in the space n-dimension to separate data with desire all of x_+ that was labeled 1, belong to positive side of super plane ($f(x_+)>0$), all of x_- were labeled -1, belong to negative side of super plane ($f(x_+)>0$). With kind of problem like SVM, a separate-data super plane is consider to be "the best" when distance from the nearest data point to super plane is biggest. And determining a document $x \notin Tr$ belong to classified c or not, follow this is determining of $f(x)$, if $f(x)>0$ then $x \in c$, else $f(x) \leq 0$ then $x \notin c$.



H3: Super plane separate training sample set

In H3, the bold line is the best super plane, and points in box are the nearest points of super plane, called support vector. And light lines which support vector stays on were called margins.

Suppose data set

$$Tr=\{(x_1,y_1),\ldots,(x_l,y_l)\}, \qquad x_i \in R^n, y_i \in \{-1,1\} \quad (4.1)$$

Case 1 If data set can be linear Separating without trouble (all of points were assigned label 1, belongs to positive side super plane and all of points were assigned label -1, belongs to negative side), we can find out a linear super plane to separate this data set like

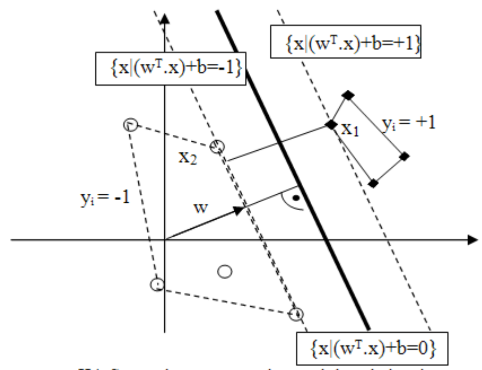
$$w^T.x+b=0 \quad (4.2)$$

With $w \in R^n$ is weight vector

$b \in R$ is free coefficient

so as to

$$f(x_i)=sign\{w^T x_i+b\}=\begin{cases}+1 & y_i=1\\ -1 & y_i=-1\end{cases} \quad \forall (x_i,y_i) \in Tr \quad (4.3)$$



H4: Super plane separate data and the relative ties

Suppose that Super plane separate data and relative ties:

$$\min_i \left|w^T.x_i+b\right|=1 \qquad i=1,\ldots,l \quad (4.4)$$

$$\text{Or } y_i\left[w^T.x_i+b\right]\geq 1, \qquad i=1,\ldots,l \quad (4.5)$$

The problem is how to determine w and b to get the best super plane, which distance from the nearest training data point to super plane is the farthest, with function of distance is

$$d(w,b;x_i)=\frac{\left|w^T.x_i+b\right|}{\|w\|} \quad (4.6)$$

$\left|w^T.x_i+b\right|$ is absolute value of $w^T.x_i+b$

$\|w\|$ is Euclid length of vector w

Suppose $h(w,b)$to be sum of distance from the nearest data point of class 1 to super plane and distance from the nearest data point of class -1 to super plane, therefore:

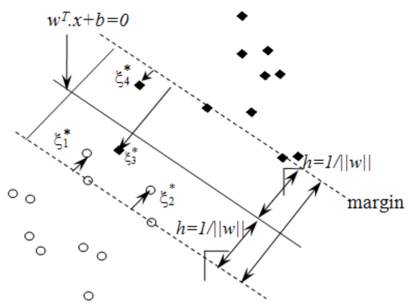
$$\begin{aligned}h(w,b)&=\min_{x_i,y_i=-1} d(w,b;x_i)+\min_{x_i,y_i=1} d(w,b;x_i)\\&=\min_{x_i,y_i=-1}\frac{\left|w^T.x_i+b\right|}{\|w\|}+\min_{x_i,y_i=1}\frac{\left|w^T.x_i+b\right|}{\|w\|}\\&=\frac{1}{\|w\|}\left(\min_{x_i,y_i=-1}\left|w^T.x_i+b\right|+\min_{x_i,y_i=1}\left|w^T.x_i+b\right|\right)\\&=\frac{2}{\|w\|}\end{aligned} \quad (4.7)$$

So, optimal super plane is a super plane with $h(w,b) = 2/\|w\|$ is maximum, equivalently with $\|w\|$ is minimum.

In summary, to find out the best super plane, we have to solve this optimal math

$$\begin{cases} \underset{w}{Min}\ \Phi(w) = \frac{1}{2}\|w\|^2 \\ y_i(w^T.x_i + b) \geq 1, \quad i = 1,...,l \end{cases} \quad (4.8)$$

Case 2: Training data set Tr can be linear Separating with interference. In this case, almost points in data set are separated by linear super plane. However, a few point are interfered, mean these points have positive label but they belong to negative side of super plane and points have negative label but belong to positive side of super plane.



H5: Case of data with interference

In this case, we replace relative tie $y_i(w^T.x_i + b) \geq 1$ by relative tie $y_i(w^T.x_i + b) \geq 1 - \xi_i \quad i = 1,...,l$ (4.9)

Here, ξ_i was called slack variable, with $\xi_i \geq 0$

$$\begin{cases} \text{Min } \Phi(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \\ y_i(w^T.x_i + b) \geq 1 - \xi_i, \quad i = 1,...,l \\ \xi_i \geq 0 \quad i = 1,...,l \end{cases} \quad (4.10)$$

C is determined parameter, define value of relative tie, the more it increase, the higher measure of violation towards experiment error.

Case 3: However, not at all of training data set can be linear Separating, in this case, we will map data vector x from n-dimension space to m-dimension space (m>n), so in this m-dimension space, data set can be linear Separating. Suppose ϕ is a *nonlinear mapping* from space R^n to space R^m.

$$\phi: \ R^n \to R^m$$

As the same time, vector x_i in space R^n will correlate with vector $\phi(x_i)$ in space R^m.

$$\begin{cases} \text{Min } \Phi(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \\ y_i(w^T.\phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1,...,l \\ \xi_i \geq 0 \quad i = 1,...,l \end{cases} \quad (4.11)$$

To calculate directly is very difficultly and complexly. If we know Kernel function $K(x_i, x_j)$, to calculate scalar product $\phi(x_i)\phi(x_j)$ in m-dimension space we needn't calculate directly with mapping $\phi(x_i)$.

$$K(x_i, x_j) = \phi(x_i)\phi(x_j) \quad (4.12)$$

And some functions often be used in document classification:

Linear function: $K(x_i, x_j) = x_i^T x_j$ (4.13)

Polynomial function : $K(x_i, x_j)=(x_ix_j+1)^d$ (4.14)

d is a natural number, from 1 to approximate10

V. Results of experimental construction

For combined CRFs and NLD

Our research group built a program to collect information from online Vietnamese news websites.
Result:

Web Sites	Sum of link	News link	Right retrieve link	Percent
The Youth	384	61	59	96%
Young people	160	67	63	94%
Zing	215	96	94	97%
News online	159	94	90	95%
Vn Express	203	98	95	96%

For SVM

Results of experimental construction for retrieved information on web pages by SVM was executed with training data set which includes 100 pieces of news on common news web pages like The Youth, Young people, Zing, News online, Vn Express.

Web sites	Sum of link	News link	Right retrieve link	Percent
The Youth	384	61	56	92%
Young people	160	67	61	91%
Zing	215	96	98	92%
News online	159	94	86	91%
Vn Express	203	98	92	94%

Although both of methods can retrieve information exactly up to over 90%, combined CRFs and NLD methods bring higher result with 97% when retrieving on Zing web site. This satisfactory result got from support of NLD method, so in the future we will try to improve SVM with support of other method to raise the accuracy.

VI. Conclusion

In this article the authors used and compared different methods for information extraction on the internet. The SVM and CRFs associated with natural language processing to solve to this problem. The experiment result of these methods were retrieved exactly up to 92% and 95% accuracy. The authors will research and inprove these methods with support of other method to raise the accuracy.

VII.References

[1] Nguyen Quang Chau, Phan Thi Tuoi, Cao Hoang Tru (2006), "Assign labels Vietnamese based on the combine of literary style and probability"

[2] Ho Tu Bao, Luong Chi Mai, "Processing Vietnamese in Information technology", page 7

[3] Dong Thi Bich Thuy, Ho Bao Quoc, "Application for Natural Language Processing on search engine in Vietnamese documents"

[4] Trang Nhat Quang (2007), "Collect information on internet to represent news on administrative web pages of city", master essay, Science and Natural University, HCMC, Vietnam

[5] Nicholas Kushmerick (1997), "Wrapper Induction for Information Extraction", *In Proceedings of IJCAI.*

[6] Hongfei Qu (2001), "Wrapper Induction: Construct wrappers automatically to extract information from web sources".

[7] Valter Crescenzi, Giansalvatore Mecca, Paolo Merialdo (2001), "RoadRunner: Towards Automatic Data Extraction from Large Web Sites", *The VLDB Journal*, pp. 109-118.

[8] John Lafferty, Andrew McCallum, Fernando Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data"

[9] Jason Weston, Support Vector Machine (and Statistical Learning Theory) Tutorial, NEC Labs America.

[10] Luong Thi Minh Hong (2006), Classified document by method *Support Vector Machine*, master essay, Hanoi, Vietnam

[11] Nguyen Linh Giang, Nguyen Manh Hien (2005), Classified Vietnamese document by SVM