

## WEBコンテンツの著作権侵害検出サーチエンジンに関する研究

研究代表者	平 澤 茂 一	早稲田大学理工学術院教授
共同研究者	児 玉 晴 男	メディア教育開発センター教授
共同研究者	常 盤 欣一朗	大阪産業大学工学部教授
共同研究者	山 崎 高 弘	大阪産業大学工学部専任講師
共同研究者	石 田 崇	早稲田大学メディアネットワークセンター助教
共同研究者	八 木 秀 樹	電気通信大学先端領域教育研究センター特任助教

### 1 はじめに

近年の情報技術の発達により、Wiki, blog などを通じて多くのユーザーが Web 上で情報を発信できる環境が整ってきている。また、検索エンジンサイトにより、キーワードを入力するだけで、関連した情報が書かれた Web ページを大量に探し出すことが可能になった。その結果、新聞記事の無断転載、Web 上の著作物文書のコピーなど著作権違反のページが増加して問題となっている。これらをすべて人手で発見するのは非常に困難である。まず、剽窃元文書を探すには類似文書を探さなければならない。高橋らの剽窃文書発見手法[3]では長さの長い単語を組み合わせて検索を行っている。しかし、長さの長い単語には外来語が多く、一般的な文書には適用が難しい。田代らの研究[6]では、隣接する文節を組み合わせて検索を行っているが、検索回数が多くなってしまふという問題点がある。[7]では文書を単語に分割し、2文書間に共起する連続単語系列を発見することによる、剽窃発見システムを提案した。しかし Web 検索を行う際のキーワードが適切でなく、あまり類似文書を得られなかった。そこで、本研究では Web 検索を行う際のキーワード選択を改良したシステムを提案する。システムを試作し、学生による剽窃レポートと著作権侵害のあった新聞記事に適用し、有効であることを示す。また、剽窃の際の言い換えに対応するため、文書内容に応じた同意語を抽出することによって剽窃文書候補を検出する手法を提案し、評価実験によりその有効性を示す。さらに、要約による著作権侵害文書を検出するためのシステムについても提案して評価を行う。

### 2 文書の著作権侵害

著作権とは、著作物の創作者である著作者に保障される権利の総称であり、知的財産権の一種である。現行の著作権法では、いくつかの条件を満たせば権利者の許諾を得ることなく文書をコピーして掲載することができる。以下にその条件を示す。

- 1) その部分を引用する必然性がある。
- 2) 引用であることが明記されている。
- 3) 著作物全体の中で自分の書いた部分が「主」、引用部分が「従」である。

以上の条件を満たしていれば、正当な引用となる。しかし、他人の文書の単なる丸写しや、「てにをは」などを少し変えただけの文書を掲載するのは無断転載あるいは剽窃となり、著作権侵害に当たる。本研究において検出対象となるのはこのような文書である。

### 3 著作権侵害 Web ページ発見支援システム

#### 3.1 システム概要

本研究では Web ページをコピーすることにより作成された著作権侵害ページを探すことを目的とし、そのようなシステムを考える。今回の研究では、太田らの研究[2]と同様に下記の3つのフェーズを用いる。

### <1> Web 検索フェーズ

対象文書が他の文書をコピーして作られたものであるか否かを調べるためには、まず類似文書を検索する必要がある。検索エンジンに問い合わせるための検索ワード生成にあたっては、文書の改変を考慮する必要がある。検索ワードは、文章の改変された部分を含まないような、検索結果が絞り込めるものがよい。そこで、複数回の検索結果の和集合を抽出することを考える。田代らの研究[6]では連続する  $k$  個の要素を検索ワードとしているが、長い文書では検索回数が非常に多くなり非効率的であると考えられる。そこで本研究では、検索ワード作成のために日本語構文解析システム (KNP[8]など) を用いて、使用する要素を決定することにする。構文解析の例を図 1 に示す。

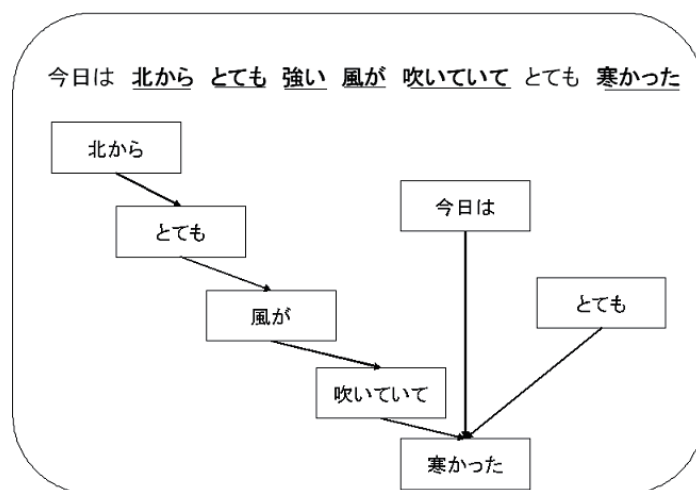


図 1：構文解析の例

独立した文節は改変される可能性が高いが、長いパスの全てが改変されてしまうことは少ない。そこで、最も長いパスに含まれる文節を利用して検索する手法を提案する。この手法によって、図 1 の例では下線のついた文節が抽出される。

#### [検索ワード生成アルゴリズム]

- (1) 元文書を文節列に分割する。
- (2) 係り受け解析を行い最も長いパスに含まれる文節を抽出する。
- (3) 連続する  $k$  個の要素を and で結合し検索ワードを生成する。
- (4)  $n-k+1$  個の検索ワードが作成されるまで、(3)を繰り返す。

検索結果の上位にランキングされている Web ページ程、剽窃者が参考に行っている可能性が高いと考えられるので、それぞれの検索ワードに対し上位  $N$  件の Web ページの URL を剽窃元候補集合として収集する。これにより、1つの Web ページに対して、複数の剽窃元候補 Web ページが得られることになる。

### <2> 剽窃ページ判定フェーズ[7]

類似文書集合を作成した後、次に対象文書と比較し、剽窃している可能性を判定する機能が必要となる。これは文書間の類似性を評価する問題とみなすことができる。文書間の類似度評価の手法としては、例えば文中の名詞と動詞を用いて文間の類似度を計算する手法[2]や、 $n$ -gram 解析により文字列の出現頻度分布を用いる手法[4]などの様々な手法が提案されている。

本研究では、Robert W. Irving[5]によって提案されたアルゴリズムを利用する。このアルゴリズムは、2文書間に一致する単語の情報をを用いて剽窃とみられる連続単語系列を発見するものである。文書  $X$  の  $i$  番目の単語  $X(i)$  と文書  $Y$  の  $j$  番目の単語  $Y(j)$  が一致することを  $X(i) = Y(j)$  と表す。文書  $X$  の  $i$  番目の単語と文書  $Y$  の  $j$  番目の単語の組におけるスコアを  $S_{ij}$  とする。

#### [Smith-Waterman アルゴリズム]

- (1) 文書  $X$  と文書  $Y$  の中から一致する単語の組を見つけ、連続単語系列の始点とする。ここでは  $X(i) = Y$

- (j) の場合を考える．初期スコアを  $S_{i,j} = 1$  とする．  
 (2) 一致した単語以降のスコアを以下のように求める．スコアが 0 になる単語の組以降のスコアは求めない．

$$S_{m,n} = \begin{cases} S_{m-1,n-1} + 1, & \text{if } X(m) = Y(n) \\ \max\{0, S_{m-1,n}, S_{m,n-1}, S_{m-1,n-1}\} - 1, & \text{otherwise.} \end{cases} \quad (1)$$

- (3) スコアを求めた範囲の一致した単語の中で，始点から最も遠い組を終点とする．

(1)から(3)で剽窃とみられる連続単語系列が 1 組得られる．これを繰り返すことで 2 文書間の全ての剽窃とみられる連続単語系列が得られる．2 つの文字列「XABCXDEXFGHXX」と「ABYCYDEFGYYYH」が与えられたとき，図 1 のようにスコアが計算され「ABCXDEXFGH」と「ABYCYDEFGYYYH」が得られる．スコアが 3 のときは互いの文書の 4 語先までの中から一致を調べる．3 語の挿入・欠落を許容するということになる．

本研究では，得られた連続単語系列内の最大スコアが 10 以上のものを剽窃とみられる連続単語系列とみなす．今回対象とする単語は，名詞と動詞のみとする．

	X	A	B	C	X	D	E	X	F	G	H	X	X
A		1	1										
B		1	2	1									
Y			1	1									
C				2	1								
Y				1	1								
D						2	1						
E						1	3	2	1				
F							2	2	3	2	1		
G							1	1	2	4	3	2	1
Y									1	3	3	2	1
Y										2	2	2	1
Y										1	1	1	1
H											2		

図 2: 剽窃発見のための Smith-Waterman アルゴリズム

Robert W. Irving の提案したアルゴリズムは単語の欠落・挿入には対応できる．しかし，日本語の剽窃に見られる文節単位や文単位の入替は考慮されていない．よって，入れ替えにより連続単語系列の長さが短くなり検出ができなくなるという問題点がある．そこで，本論文では我々が以前に提案した手法[7]を用いる．本研究では検出する連続単語系列の単位を 3 単語とした．そして，「AXB」と「BYA」となった場合「X」と「Y」の部分がともに 5 語以内の場合に結合するものとした．

### <3> 検査者提示フェーズ

剽窃している疑いが高い対象文書とその剽窃元と思われる文書が見つかったとしても，最終的に剽窃か否かを判断するのは人手にまかされることになる．そのため，剽窃か否かをできるだけ容易に判断できるようにするために，それらを効果的に表示する機能が必要となる．本研究では，チェック対象ページと剽窃元候補のページの連続一致単語系列が含まれる部分を文書中から抽出し剽窃検査者に表示することによりこれを実現する．

## 3.2 評価実験

本研究で述べたシステムを実装し，評価実験を行った．本実験では，検索フェーズには Web ページの剽窃の含まれる学生レポートを用いる．判定フェーズには著作権侵害のあった新聞記事を用いる．形態素解析には茶筌[9]を用いた．パラメータは  $N = 20, k = 2$  とした．本システムでは，Yahoo! Japan[1]の提供するサービスを利用している．なお，検索の制限回数は 50000 回/24h，1 検索あたりの最大検索結果数は 50 である．

### 3.2.1 学生レポートを用いた評価実験

本実験では Web ページの剽窃の含まれると思われる学生レポートから剽窃元の Web ページを検索する．

課題：情報化社会とアウトソーシングについてレポート数：20 件

科目名：情報化社会概論

レポートは電子メールで提出され、形態素解析には茶筌を用いた。

### (1) 検索フェーズ

検索フェーズでは全ての文節を用いる手法と本手法の比較を行った。その結果を表1に示す。我々の提案により検索回数、計算量などを削減しながらも多くの剽窃候補が得られた。これは提案手法により不要な文節が取り除かれて、剽窃の検索に適した文節がキーワードとして選択されたためである。

1つのレポートから多くの剽窃が発見された理由は、有名なWebページを他のWebページが剽窃しているためである。剽窃文書から他の剽窃文書が発見できていることから検索フェーズとして良い性能を示しているといえる。

### (2) 検索フェーズ

提案手法では従来手法より多くの文書が剽窃元文書と判定された。他のWebページを剽窃したWebページを多く検出しているため、剽窃判定数がかかなり多くなった。実際に剽窃があったかはわからないが、我々が人手で確認した結果、全ての剽窃元文書が妥当であると判断された。これは、学生レポートを作る際に、文書の改変をあまり行わなかったため検出が容易であったためだと思われる。

#### 3.2.2 新聞記事を用いた追加実験

学生レポートは実際に剽窃をしたのかが確認できないため、剽窃の判明している新聞記事で剽窃判定フェーズの追加実験を行う。実験データには著作権侵害のあった山梨日日新聞の新聞記事を用いる。社説の盗用があったと発表されている文書は対照表として公開されており、その中の文書を対象として実験を行った。これらの新聞記事には改変が多く、これらの判定ができれば一般的な文書十分な結果が期待できる。今回対象とした剽窃箇所は51カ所、そのうち剽窃と判断されたのは44カ所であった。実際に剽窃部分と判断された部分を図2に、対照表の中で検出できなかった部分を図3に示す。

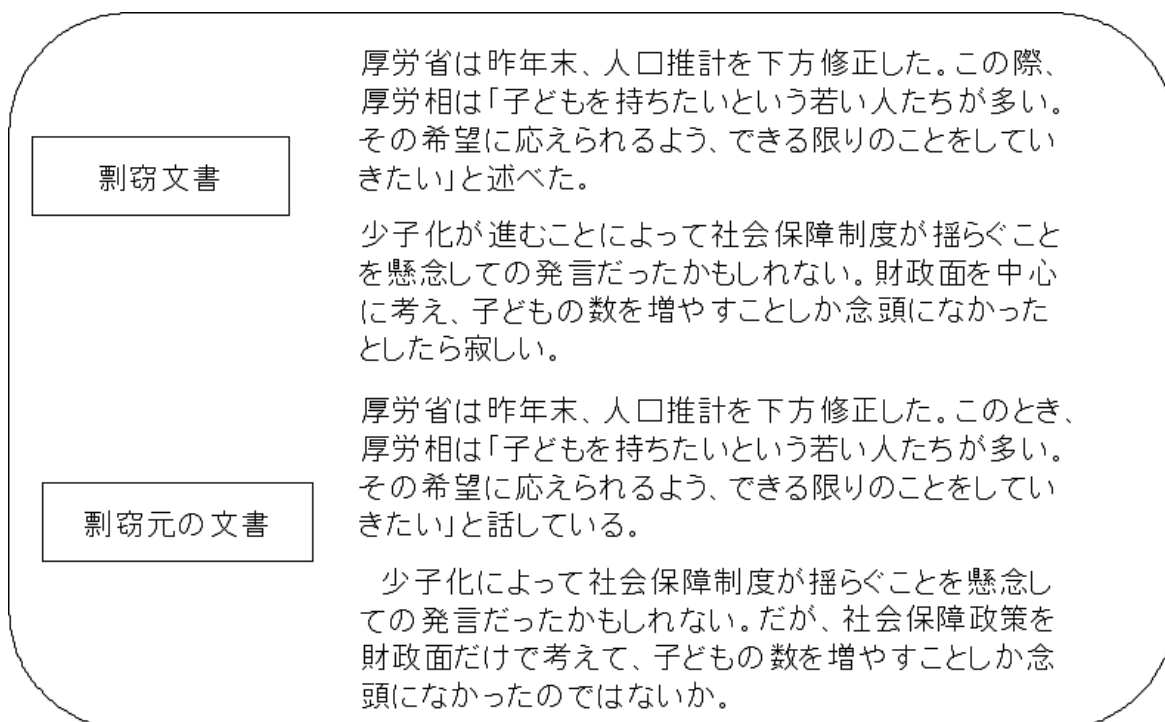


図3：剽窃と判断された文書の例

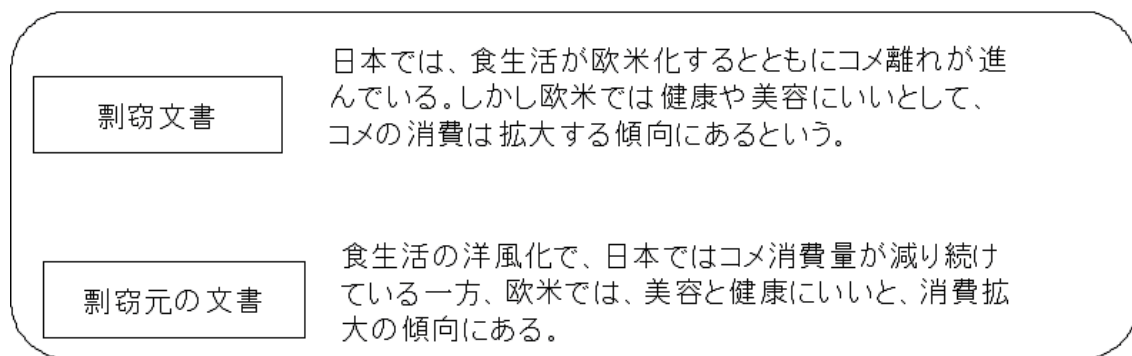


図 4: 剽窃と判断されなかった文書の例

剽窃と判断されなかった部分は、言い換えや入れ替えが多く今回のシステムでは検出することができなかった。また、対照表には無いが検出された文章があった。事実を述べている文章が多く新聞社は剽窃と判断しなかったものと思われる。人手で調査しても見解が分かれるところであり、システムの動作としては問題ないと考えている。

#### 4 同意語抽出による著作権侵害候補文書発見手法

前章では、著作権侵害ページを自動的に検出するための剽窃 Web ページ発見システムの構築を行った。剽窃文書を検出するためには、まず手元の著作権保有文書をもとにして、Web 上から著作権を侵害していると思われる文書をできる限り多く集めなければならない。しかしながら、剽窃者は自分が剽窃を行ったことが容易にはわからないように、語句の言い換えや語尾の変化、順序の入れ替えなどの編集を行う。そのため、収集の際には様々な検索クエリを考えて試行錯誤しなければならず、中でも語句を言い換えている剽窃文書は、検索クエリの与え方によっては、全く検出されない可能性がある。

この問題を解決するためには、著作権文書中の語句の同意語で検索する必要がある。Web 文書の語に対して同意語を抽出する方法としては、検索語の共起情報を用いたクラスタリング手法[10]、閲覧履歴に基づいたシソーラスの構築手法[11]、限定的同意語を利用したアンケート結果を分類手法[12]などがある。本研究では、特定的话题に対して共起する同意語を抽出することで、著作権侵害候補文書を発見する手法を提案する。

著作権侵害の対象をWeb上に存在するページ（文書）とし、剽窃Webページを発見するシステムを考える。対象システムは前章と同様

- (1) Web検索フェーズ
- (2) 剽窃ページ判定フェーズ
- (3) 検査者提示フェーズ

からなる。

検索フェーズ（図5）では、一般の検索エンジンを用いてキーワード検索を実行し、文書を収集する。ここでは、「検索クエリをどのように作成するか」が問題となる。これまでは、

- ・著作権文書の中で連続する単語をクエリとする。単語の組を順番にずらして網羅的に検索し、文書を集める[6]。
- ・著作権文書の構文解析を行い、語句の係り受け関係を考慮して、文中で主となる要素を判定し、その語を検索クエリとする（前章）。

といった方法を用いてきた。しかしながら、語句が改変された場合、剽窃文書が検索結果に出力されない可能性がある。手元の文書中の語句から作成したクエリでは、言い換えのある侵害文書をWeb上から見つけてくるのが難しい。そこで、語句の改変に対応した収集方法を導入して、著作権侵害候補の文書集合を集めるようにする。最終的には、保有する著作権文書に対して、できるだけ多くの著作権侵害Webページの候補集合を収集する検索サーバ、クローラーを構成することが目的である。



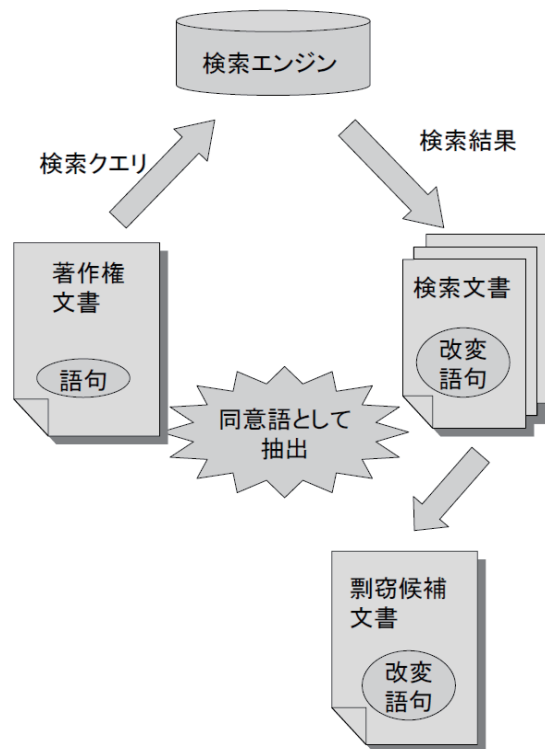


図5：Web 検索フェーズ

#### 4.1 Web 文書での同意語

自由記述のWeb 文書の場合、記述に制約がないために、同様の内容について様々な言い回しで表現される。そのため、一般的には同意語や類義語とならない語句についても、特定の内容のもとでは、同じ事を意味している場合がある。また、検索エンジンのクエリには固有名詞や新語が指定されることが多いという報告もある。そのため、シソーラス等による語句の単純な読み替えでは、同意語抽出の対応ができないといえる。本研究で対象とする剽窃文書については、言い回しの変化が故意になされていると考えると、同等の問題であるとみなすことができる。既存システムにおいて、剽窃文書と判断されなかった例を以下に示す。この記事は剽窃であったことが発表されている。

- ・元記事

「食生活の洋風化で、日本ではコメ消費量が減り続けている一方、欧米では、美容と健康にいいと、消費拡大の傾向にある。」

- ・剽窃記事

「日本では、食生活が欧米化するとともにコメ離れが進んでいる。しかし欧米では健康や美容にいいとして、コメの消費は拡大する傾向にあるという。」

『食生活』という内容においては、「欧米」と「洋風」は同意語である。しかしながら、一般的には同意語とはいえない。そのため、シソーラスを用いるのではなく、文書内容に応じて同意語を抽出し、それにもとづいて剽窃文書を検索する必要がある。

#### 4.2 共起の類似性による同意語の抽出

文書中の語句をクラスタリングする研究は数多くなされている。語句の類似性に注目する方法では、語 $w_1$ と語 $w_2$ について、文書中の他の語句との共起分布を調べる。共起分布が似ていれば、語 $w_1$ と語 $w_2$ を同じクラスタに分類する。この考えをもとに語句の共起状況により同意語候補を抽出し、剽窃文書の検索に利用する手法を提案する。さらに、同意語は同じ意味を持ち、同じ使われ方をするため、置き換えが可能であると考えられる。そこで、同意語の検証は、同意語候補の組について、他に言い換えた文書が実際に存在しているかどうかで判断するアプローチをとる。語 $w_1$ と語 $w_2$ が同意語の候補かどうかは、他の語句との共起分布を比較することで定まる。本研究では共起度の尺度として、相互情報量 $MI$  (Mutual Information)を用いる。

検索・収集した文書集合の中に、語 $w_1$  および語 $w_2$  とは異なる語句が $A_1, \dots, A_n$  の $n$  個あったとする。語 $w_1$  と語 $A_i$  の共起度および語 $w_2$  と $A_i$  の共起度は、

$$MI(w_1, A_i) = \log \frac{P(w_1, A_i)}{P(w_1)P(A_i)} = \log \frac{N \cdot freq(w_1, A_i)}{freq(w_1)freq(A_i)}, \quad (i=1,2,\dots,n), \quad (2)$$

$$MI(w_2, A_i) = \log \frac{P(w_2, A_i)}{P(w_2)P(A_i)} = \log \frac{N \cdot freq(w_2, A_i)}{freq(w_2)freq(A_i)}, \quad (i=1,2,\dots,n), \quad (3)$$

で求める。ここで、 $N$  は語の総数、 $freq(x)$  は語 $x$  の出現頻度、 $freq(x, y)$  は語 $x$  と語 $y$  が共起して出現する頻度である。

語 $w_1$  と $A_1, \dots, A_n$  の共起分布を共起特徴ベクトル $w_1$  とし、同じく語 $w_2$  と $A_1, \dots, A_n$  の共起分布を共起特徴ベクトル $w_2$  で表す。

$$w_1 = (MI(w_1, A_1), MI(w_1, A_2), \dots, MI(w_1, A_n)) \quad (4)$$

$$w_2 = (MI(w_2, A_1), MI(w_2, A_2), \dots, MI(w_2, A_n)) \quad (5)$$

$w_1$  と $w_2$  の余弦値 $\cos(w_1, w_2)$  を語 $w_1$  と語 $w_2$  の類似度 $sim(w_1, w_2)$  とする。類似度が一定値以上の場合、それらを同意語候補として記憶しておく。

$$sim(w_1, w_2) = \frac{w_1 \cdot w_2}{\|w_1\| \cdot \|w_2\|} \quad (6)$$

語 $w_1$  と語 $w_2$  が同意語として利用できるかどうかは、語を入れ替えた文書が実際に存在するかどうかで判断を行う。

### 4.3 評価実験

図6 に提案手法による剽窃候補文書の検索手法の流れを示す。前処理として、著作権文書および収集文書を形態素解析によって単語に分割する。次に語句の集合に対し、語句間の共起度を調べる。共起度をもとに、同意語の候補を選び、候補となった単語の組について、フィードバックさせ検証し、同意語を抽出する。抽出された同意語を用いて、単語の比較による同意語判定を行う。著作権侵害検索システムの場合、対象となる文書をWeb上から集めてくる必要がある。本研究の場合、まずは手元の著作権文書中の単語クエリから検索を開始し、徐々に同意語として使われる単語を増やし、その単語を利用して再検索を繰り返すという方針をとる。具体的な処理については、以下に示す。

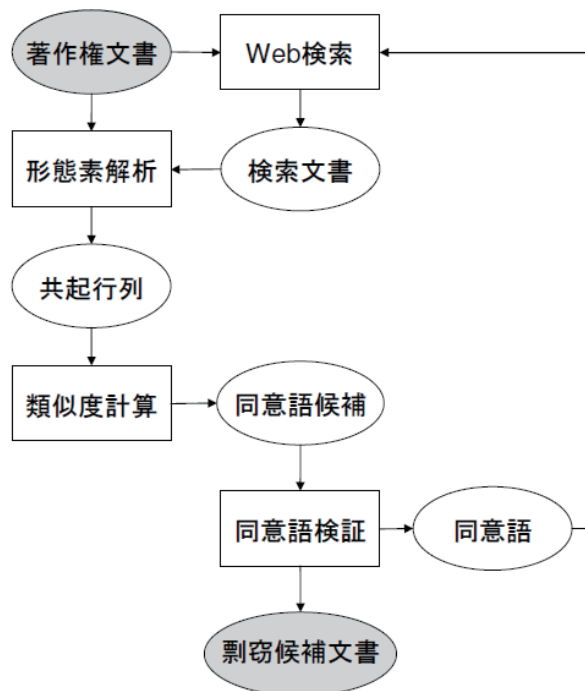


図6：剽窃文書検索処理手順

### (1) 形態素解析による前処理

著作権文書および検索エンジンからの収集文書に対して形態素解析を行い、単語の品詞情報をもとに対象となる語句を選択する。本研究では日本語形態素解析システム「茶筌」[9]を使用して、単語に分割し、品詞、読み、活用形などの情報を単語に付与した。単語の品詞の情報をもとに、以下のような処理により使用する語句を選択する。

- ・ 肯定／否定を判断  
否定を意味する助動詞を検出し、直前の語句を否定形と判断する。
- ・ 不要語の排除  
「する」や「ある」などの同意語判定には不要な語句を削除する。
- ・ 品詞の選別  
名詞…数、接尾、代名詞、非自立、副詞化を削除  
形容詞…自立以外は削除  
動詞…自立以外は削除  
副詞…すべて含む

### (2) 共起行列を作成

語句の共起の頻度を集計し、表1のような共起行列を作成する。同意語は同じ品詞で判断をするため、「名詞」「動詞」「形容詞」といった品詞ごとに共起行列を求める。表1の場合、『食生活』という語句と『良い』という語句は、13文で共起していることを示している。

表1:共起行列

	日本	食生活	欧米	健康	美容	消費	利用	人気	洋風	…
食べる	15	10	4	8	2	11	3	2	7	
おいしい	5	8	3	4	2	4	2	3	10	
良い	2	13	5	18	8	2	1	4	2	
あう	12	2	9	0	3	5	0	5	3	
増える	3	5	8	0	3	6	1	10	4	
食べる(否)	2	5	3	3	0	2	0	1	0	
悪い	4	2	10	6	9	7	2	6	3	
…										

### (3) 類似度の計算

共起行列から、ある語句に対しての他の語句の共起度を求め、それらの分布を特徴ベクトルとする。任意の語句の特徴ベクトルに対して、ユークリッド類似度を求め、閾値（ここでは0.6とした）以上であれば、その語句の組は同意語候補として抽出する。

### (4) 同意語候補の検証

同意語候補の組を実際の文書にフィードバックさせて検証を行う。収集文書中に語句を置き換えた文書が存在すれば、同意語として情報を保存する。さらに、収集文書を剽窃候補文書として保管しておく。

### (5) Web 文書の再検索

新たに抽出した同意語を新たな検索クエリとして、検索エンジンを用いてWeb上から文書の再検索を行う。これらの処理を何度か繰り返すことで、多くの著作権侵害が疑われる候補文書を集めることができた。今回の評価実験では、一つの著作権文書をもとに剽窃候補文書を探索したところ、以下のような結果が得られた。

- ・ 抽出同意語…12語
- ・ 剽窃候補文書…118件

剽窃候補文書を調査したところ、語句の置き換えにより作成されている文書が確認された。しかしながら、これらが本当に剽窃文書であるかどうかは、最終的には専門家に判断を委ねることになる。



## 5 おわりに

本研究では、本研究では Web 検索を行う際に

1. キーワード選択を改良したシステム
2. 同意語抽出による改良システム

を提案した。その結果、本システムが効率的に剽窃を検出できることを示した。特に、1 では検索フェーズにおいて計算量の削減しながら多くの剽窃候補ページを収集することに成功した。これにより、違反が多いホストの管理者や作成者個人に警告を行うことができる。また、著作権侵害の抑制にも役立つと思われる。さらに、2 では剽窃の際の書き換えにも対応することが可能となった。

今後の課題は、剽窃ページ判定フェーズにも今回の同意語判定の手法を組み込むことで判定精度を上げ、自動的に著作権侵害を判定するシステムを作成することである。また、シソーラスを活用することによって同意語や多義語へ対応することである。さらには、検索サーバ、クローラーを作成して自動的に著作権侵害の可能性がある文書を収集するシステムの構築が考えられる。

## 【参考文献】

- [1] Yahoo! JAPAN, <http://www.yahoo.co.jp/>
- [2] 太田貴久, 増山繁, “学生レポート採点支援のためのレポート類似部分発見手法”, 信学技報, NLC2005-112, pp.37-42, 2006.
- [3] 高橋勇, 宮川勝年, 小高知宏, 白井治彦, 黒岩丈介, 小倉久和, “WEB からの剽窃レポート検出手法の実装と評価”, 人工知能学会研究会資料, SIG-ALST-A503-01, 2000.
- [4] 深谷亮, 山村毅, 竹内義則, 松本哲也, 工藤博章, 大西昇, “単語の頻度統計を用いた文章の類似性の定量化”, 電子情報通信学会論文誌, J87-D-II, 02, pp.661-672, 2004.
- [5] R.W. Irving, “Plagiarism and collusion detection using the smith-waterman algorithm,” Technical Report 164, Dept of Computing Science, University of Glasgow, 2004.
- [6] 田代崇, 上田高德, 堀泰祐, 平手勇宇, 山名早人, “Web 上の文章を対象とした著作権違反自動検知システム”, 日本データベース学会 Letters Vol.5, No.2, 2006.
- [7] 高島秀佳, 坂口朋章, 長尾壮史, 石田崇, 平澤茂一, “著作権侵害検出を目的とした類似文書発見手法”, 経営情報学会, 2006 年度秋季全国研究発表大会予稿集, pp.58-61, 2006.
- [8] 日本語構文解析システム KNP, <http://nlp.kuee.kyoto-u.ac.jp/nlresource/knp.html>
- [9] 形態素解析システム茶筌, <http://chasen.naist.jp/hiki/ChaSen/>
- [10] 有田一平, 菊池英明, 白井克彦, “検索語の共起情報を利用した単語クラスタリングと Web 検索への応用,” 情報処理学会研究報告自然言語処理研究会報告, Vol.2007, No.76, pp.115-120, 2007.
- [11] 安川美智子, 山田篤, “Web 検索エンジンを用いた用語検索履歴からのシソーラス自動構築手法の評価と改良,” DEWS2005 論文集, 5C-i8, 2005.
- [12] 国定美佐, 平松綾子, 能勢和夫, “自由記述アンケート分類のための限定的同意語特定手法,” 第 50 回自動制御連合講演会予稿集, pp.367-370, 2007.
- [13] 中山信弘, マルチメディアと著作権, 岩波書店, 1996.

## 〈発表資料〉

題名	掲載誌・学会名等	発表年月
著作権侵害検出のための剽窃 Web ページ発見システム	経営情報学会, 2007 年度秋季全国研究発表大会予稿集	2007 年 11 月
著作権侵害文書検出のための要約文発見手法	経営情報学会, 2007 年度秋季全国研究発表大会予稿集	2007 年 11 月
著作権侵害検出を目的とした同意語抽出による候補文書発見手法	経営情報学会, 2008 年度春季全国研究発表大会予稿集	2008 年 6 月