

適合性と具体性を考慮したキーワードのランキング

Ranking Keywords Based on Relevance and Concreteness

吉田 大我†
Taiga Yoshida入江 豪†
Go Irie佐藤 隆†
Takashi Satou小島 明†
Akira Kojima

表1 YouTubeの映像におけるタグの例

roger federer novak djokovic us u.s. open 2009 tennis spectacular shot amazing phenomenal sport sports ny new york flushing meadows point match crazy

1. はじめに

ユーザがアクセス可能なオンライン映像は、今もなお劇的な増加を続けている。動画共有サイト YouTube^(注1)では、2010年3月時点で、世界中のユーザによって毎分24時間に相当する量の映像が投稿されていると言われている。オンラインでの映像配信は、ウェブ上のみにとどまるものではない。IPTV (Internet Protocol Television) や VOD (Video On Demand) などのテレビ番組配信や映像コンテンツ提供サービスも普及したことにより、ユーザは簡単に多くの映像を視聴することができるようになってきている。このような爆発的な映像の増加は、ユーザに豊富な選択肢を与える一方で、膨大な映像の中から、視聴したい映像を探し出すという大きな困難を生みだしている。

現在、普及している動画共有サイトや IPTV および VOD によるテレビ番組の配信では、検索クエリとしてキーワードを受け付けている。これらのサービスでは、入力されたキーワードと、映像に付与されたメタデータを比較し、マッチしたものを検索結果としてリスト表示する。メタデータには、映像のタイトルや概要文、あるいは、タグなどのキーワードが含まれる。しかしながら、単純なキーワードのマッチングによる検索では、精度の高い検索結果を得ることは難しい。例えば、表1は、YouTubeにおいて“Federer Amazing Shot at the US Open 2009 semifinal”というタイトルの映像に付与されたタグの一覧である^(注2)。この映像において、“sport”や“tennis”などのタグは映像と強く関連していることが容易に推測できるが、“match”や“2009”といったタグは、必ずしも映像の主要な内容とは関連していない。単純なキーワードのマッチングによる検索では、タグが内容と強く関連しているかどうかを考慮しないため、内容とあまり関係しない映像を検索結果リストに表示してしまう可能性がある。

このような問題を解決するためには、どのキーワードが映像の内容に対して重要なものであるかを推定し、マッチングの際に考慮することが必要となる。最近、情報検索やメディア処理の分野では、コンテンツに付与されたキーワードの中から、内容と適合するキーワードを推定する試みが報告されている [1, 2]。これらの手法では、キーワード同士の共起性や、キーワードとコンテンツの特徴量との共起性を分析することにより、コンテンツの内容とキーワードの適合性を推定している。

しかしながら、内容に適合したキーワードを推定するのみでは、十分な検索結果が得られるとは言い難い。例えば、表1に挙げた映像の例において、“sport”や“tennis”はいずれも映像の内容と適合したキーワードであるが、両者を比較すると、“sport”より“tennis”の方が映像の内容を具体的に表現したキーワードであると考えられる。

“sport”というキーワードは、“tennis”だけではなく、“baseball”や“football”などにも関連する、より抽象的なキーワードである。したがって、映像の内容に対して得られる情報としては、“sport”から得られる情報よりも、“tennis”から得られる情報の方が多いため、本研究の着眼点は、1) 内容に適合している、さらに、2) 内容をより具体的に表現するキーワードの方が、より重要なキーワードであるという考えである。

本研究では、従来着目されてきた内容との「適合性」に加えて、さらに、どのくらい内容を絞り込むことができるかを表す「具体性」の観点からキーワードのランキング手法を提案する。本手法は、映像に付与されたキーワード間の関係を分析することにより、内容との適合性を推定し、キーワードの共起性のばらつきを分析することによって具体性を推定する。これらの2つの推定結果を組み合わせることにより、内容に適合し、かつ具体性も高いキーワードから順にランキングして提示する。本報告では、約300,000本のIPTVの番組情報および約50,000本のYouTubeの映像情報から推定したキーワードの適合性、具体性の結果を用いた評価実験結果を報告する。

2. 関連研究

画像や映像などのコンテンツに対して、重要なキーワードを抽出する、あるいは、付与するといった問題には、さまざまな取り組みがなされてきている。代表的なものとしては、コンテンツの内容に合うタグを自動付与するオートタギングや、付与されているキーワードが内容と適合しているかを判定する手法に対する取り組みがある。

2.1 オートタギング

オートタギングの基本的なアプローチの一つは、内容の類似したコンテンツに既に付与されているタグを推薦するものである。コンテンツが画像や映像の場合、コンテンツの内容は画像特徴量によって表現できる。そのため、画像特徴量が類似するコンテンツのタグを推薦する手法が盛んに研究されている [3-5]。しかしながら、画像特徴量を利用したオートタギングは精度の向上が難しく、計算時間もかかるため、まだ実用には至っていない。一方、画像特徴量を利用せず、既に付与されているタグと共起しやすいタグを推薦するアプローチも検討されている [6, 7]。しかしながら、オートタギングは、コンテンツに新しいタグを付与

(注1) YouTube, <http://www.youtube.com/>(注2) <http://www.youtube.com/watch?v=RJuEzJEQ9N4>

†日本電信電話株式会社 NTT サイバーソリューション研究所

することを目的しているものであり、既に付与されているタグに対する評価を目的としているわけではない。本研究では、既に付与されているキーワードに対して、内容との適合性・具体性を評価する手法を提案する。

2.2 内容に適合したキーワードの推定

より最近では、既に付与されているタグやキーワードのうち、コンテンツの内容に適合しているものを自動判定する試みが報告されてきている。例として、Liu らによるタグランキング手法 [1] や Iwata らによる適合性判定手法 [2] が挙げられる。Liu らは、タグによる画像検索の精度向上や、画像に付与すべきタグの推薦を目的として、画像に付与されたタグを、画像との適合性が高い順にランキングするタグランキングを提案した。Iwata らは、トピックモデルを利用し、教師なしで内容に適合するタグを推定する手法を提案した。これらの手法は、いずれも内容と適合するキーワードであるかどうかという点は考慮しているものの、話題を絞り込むことができるキーワードであるかどうかという「具体性」の観点には着目していなかった。映像の内容をより具体的に表現するキーワードを抽出するため、キーワードの具体性に注目した点が、本研究の大きな特徴である。

3 適合性と具体性によるキーワードランキング

提案手法では、以下の2つの指標において、双方とも高い値をとるようなキーワードを上位にランキングする。

- キーワードの適合性：どれだけ映像内容と関連したキーワードであるか
- キーワードの具体性：どれだけ話題を絞り込むことができるキーワードであるか

提案手法の流れを図1に示す。ただし、 c_n は全N個の映像コンテンツのうちの一つであり、 $k_m^{c_n}$ は c_n に付与された M_{c_n} 個のキーワード集合 $K^{c_n} = \{k_m^{c_n} | m = 1, 2, \dots, M_{c_n}\}$ のうちの一つである。キーワードと内容との適合性の計算では、まず、映像コンテンツ c_n に付与されたキーワード $k_m^{c_n} \in K^{c_n}$ のうち、任意の2つのキーワード間の関連の強さを算出する。次に、この関連の強さに基づいて $k_m^{c_n}$ の、 c_n の内容との適合性を表す値 $S_r(c_n, k_m^{c_n})$ を算出する。キーワードの具体性の計算では、映像集合全体において、各キーワード $k_m^{c_n}$ と共起するキーワードの種類の数ばらつきを分析することにより、キーワードの具体性を表す値 $S_c(k_m^{c_n})$ を算出する。最後に、適合性と具体性の双方が高いキーワードをランキングするための評価関数 $S(c_n, k_m^{c_n})$ を、以下の式(1)により定義する。

$$S(c_n, k_m^{c_n}) = S_r(c_n, k_m^{c_n}) \times S_c(k_m^{c_n}) \quad (1)$$

ある映像コンテンツ c_n と、 c_n に付与されたキーワード $k_m^{c_n}$ があったとき、 $k_m^{c_n}$ が c_n の内容によく適合するキーワードであった場合、 $S_r(c_n, k_m^{c_n})$ の値は大きくなる。また、 $k_m^{c_n}$ が話題を絞り込むことができるようなキーワードであった場合、 $S_c(k_m^{c_n})$ の値は大きくなる。したがって、 $k_m^{c_n}$ が c_n の内容と適合し、同時に $k_m^{c_n}$ が具体的なキーワードであった場合に、 $S(c_n, k_m^{c_n})$ の値は大きくなる。そこで、提案手法では、 $S(c_n, k_m^{c_n})$ の降順に、キーワードをランキングして出力する。

3.1 適合性の計算

まず、キーワードと内容との適合性の計算方法について説明する。これまでの研究では、キーワードの適合性を求

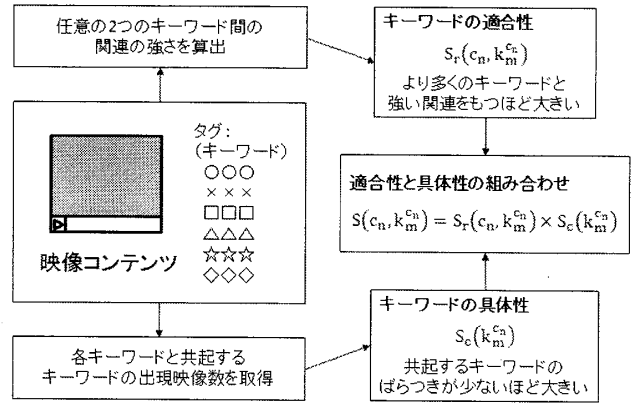


図1 キーワードの適合性および具体性の算出手順

める際には、キーワードの共起関係を利用することが一般的である [1, 3]。提案手法においても、キーワードの共起関係を用いてキーワードと内容との適合性を算出する。いま、内容と適合するキーワードが持つ特徴を考えると、映像の主要な内容と関連するキーワードは、内容と適合していると言えるであろう。また、映像の主要な内容に関する話題は、より多くのキーワードによって重複して表現されると考えられる。そこで、本研究では、同一コンテンツに付与された他のキーワードと、どれだけ同じ内容を表現できているかによりキーワードの適合性を測ることを試みる。

まず前処理として、意味を持たないキーワードを除外する。Wu ら [3] によると、コンテンツに付与されるキーワードのうち、出現回数の高い上位約 1%は、一般的すぎてほとんど意味を持たないキーワードであるとされる。そこで、まずコンテンツに付与された全てのキーワードのうち、「番組」、「放送」、「情報」など、出現頻度の高い上位約 1%のキーワードを適合性の算出対象から除外した。

続いて、キーワード間の共起性を算出する。映像コンテンツ c_n に付与されたキーワード集合 K^{c_n} に含まれる任意の2キーワード k_x, k_y 間の関連の強さ $R(k_x, k_y)$ を、以下の式(2)および(3)により定義する。

$$R(k_x, k_y) = \begin{cases} \frac{N \times (N_{k_x, k_y} \times N_{\bar{k}_x, \bar{k}_y} - N_{\bar{k}_x, k_y} \times N_{k_x, \bar{k}_y})^2}{N_{k_x} \times N_{\bar{k}_x} \times N_{k_y} \times N_{\bar{k}_y}} & (2) \\ \frac{N \times (N_{k_x, k_y} \times N_{\bar{k}_x, \bar{k}_y} - N_{\bar{k}_x, k_y} \times N_{k_x, \bar{k}_y})^2}{N_{k_x} \times N_{\bar{k}_x} \times N_{k_y} \times N_{\bar{k}_y}} & (3) \end{cases}$$

ただし、 $N_{k_x}, N_{\bar{k}_x}, N_{k_x, k_y}, N_{\bar{k}_x, \bar{k}_y}, N_{\bar{k}_x, k_y}, N_{k_x, \bar{k}_y}$ はそれぞれ、 k_x が付与された、 k_x が付与されていない、 k_x と k_y を共に付与された、 k_x と k_y のどちらも付与されていない、 k_x は付与されず k_y は付与されている、 k_x は付与され k_y は付与されていない映像の総数である。また、 $N_{k_x, k_y} / N_{k_x, \bar{k}_y} \geq N_{\bar{k}_x, k_y} / N_{\bar{k}_x, \bar{k}_y}$ のときは(2)、それ以外の場合は(3)とする。 $R(k_x, k_y)$ は、映像集合における各キーワードの出現傾向が独立であることを帰無仮説とするカイ 2乗検定統計量の値に相当する。キーワード k_x と k_y の出現頻度に相関がない場合は0に近い値を取る。相関がある場合は正の相関であれば正の値、負の相関であれば負の値を取り、相関が高いほど絶対値は大きくなる。そのため、 $R(k_x, k_y)$ により、各キーワード間の関連の強さを測ることができる。

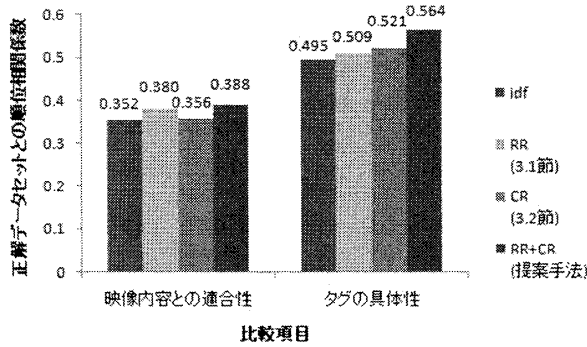


図2 各手法と適合性・具体性の順位相関係数(YouTube)

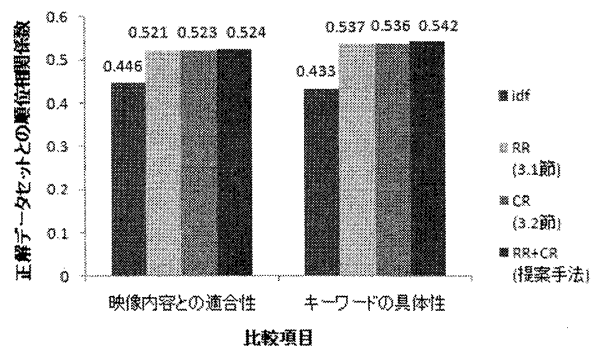


図3 各手法と適合性・具体性の順位相関係数(IPTV)

最後に、キーワード間の関連性に基づいて、キーワードの内容に対する適合性を求める。キーワード $k_m^{c_n}$ がどれだけコンテンツ c_n の内容に適合しているかを表す値 $S_r(c_n, k_m^{c_n})$ を以下の式(4)によって定義する。

$$S_r(c_n, k_m^{c_n}) = \sum_{k_j \in K^{c_n} \text{ s.t. } k_j \neq k_m^{c_n}} R(k_j, k_m^{c_n}) \quad (4)$$

キーワード $k_m^{c_n}$ が他の多くのキーワードと強く関連しているほど、 $S_r(c_n, k_m^{c_n})$ の値は大きくなる。したがって、 $S_r(c_n, k_m^{c_n})$ を計算することが、キーワードがどれくらいコンテンツの内容に適合しているかを求めるための指標になると考えられる。

3.2 具体性の計算

次に、キーワードの具体性の計算方法について説明する。いま、具体性のあるキーワードが持つ特徴を考えると、キーワードを見ることで話題を推測可能なキーワードは、具体的であると言えるであろう。また、共起するキーワードのばらつきが少ないとき、他のキーワードを推測しやすくなる。つまり、話題を推測しやすくなると考えられる。そこで、本研究では、ばらつき度合いの指標であるエントロピーを元に、キーワードの具体性を算出することを試みる。

映像コンテンツ c_n において、キーワード $k_m^{c_n}$ の具体性がどの程度であるかを算出する場合を考える。いま、全映像コンテンツにおいて、キーワード $k_m^{c_n}$ と共起するキーワードの集合 T に対し、 T に含まれる各キーワードが付与された映像数を要素とするベクトルのエントロピー $E(T)$ を、以下の式(5)により求める。

$$E(T) = - \sum_{k_i \in T} \frac{N_{k_i}}{N_T} \log \frac{N_{k_i}}{N_T} \quad (5)$$

ただし、 N_{k_i} はキーワード k_i が付与されている映像の総数、 N_T は各キーワードが付与された映像数の総和であり、 $N_T = \sum_{k \in T} N_k$ である。いま、 $k_m^{c_n}$ と共起するキーワードの種類のばらつきが小さければ $E(T)$ は大きくなり、ばらつきが大きければ $E(T)$ は小さくなる。そこで、キーワードが具体的であるかどうかを表す値 $S_c(k_m^{c_n})$ を、以下の式(6)によって定義する。

$$S_c(k_m^{c_n}) = \sigma^{-E(T)} \quad (6)$$

σ はパラメータであり、本研究では、 $\sigma = 2$ とおいた。また、 $S_c(k_m^{c_n})$ は、キーワードの具体性が高いときに大きくなるよう、 $E(T)$ の正負を入れ替えてある。 $S_c(k_m^{c_n})$ は、キーワード $k_m^{c_n}$ が付与されている映像の集合において、 $k_m^{c_n}$ 以外に

も同じキーワードが付与されていることが多い場合、すなわち同じような話題の映像である可能性が高い場合に値が大きくなる。そのため、 $S_c(k_m^{c_n})$ の値を調べることにより、各キーワードが具体性のあるキーワードであるかどうかを判定することができる。

4. 評価実験

3.1節および3.2節で示した手法によって、それぞれキーワードの適合性と具体性によるランキングが行えることを確かめるため、評価実験を行った。実験には、315,931本のIPTVの番組情報および50,013本のYouTubeの映像に付与されたキーワードの情報を格納したデータベースを利用した。実験では、式(4)のキーワードの適合性によるランキング、および式(6)の具体性によるランキングを作成し、それらを組み合わせた式(1)によるランキングも作成し、人手によるランキングとの比較を行った。

4.1 正解データの取得

本実験では、YouTubeおよびIPTVの映像から、それぞれ5個以上20個以下のキーワードが付与されている映像をランダムに50本を選び、正解データの作成を行った。実験では、IPTV、YouTubeの映像について各6名、計12名の被験者に協力を依頼した。実験では、1人につきYouTubeもしくはIPTVの映像25本について、それらに付与されたキーワードが、映像内容との適合性の高いキーワードであるかどうか、話題を絞り込むことができる具体的なキーワードであるかどうかを5段階で評価する形式で行った。各映像に付加されたキーワードは3人の評価者によって評価されるものとし、各映像に付与されたキーワード集合を、適合性もしくは具体性の評価の平均値が高い順に並べたものを正解データとした。

得られた正解データをもとに、各評価者によるランキング間でのスピアマンの順位相関係数の平均を計算したところ、キーワードの適合性については約0.556(YouTube)と約0.615(IPTV)、キーワードの具体性では約0.588(YouTube)と約0.667(IPTV)となった。

4.2 評価方法

評価は、提案手法によって並べ替えたキーワードのランキングと、正解データにおける適合性もしくは具体性のランキングとのスピアマンの順位相関係数を求めることによって行った。順位相関係数は-1から1までの値をとり、値が大きいくほど正解データとの相関が高いことを意味する。

本実験では、提案手法である式(1)によるランキング(RR+CR)の他に、以下の3つの手法を比較する。

idf (ベースライン) : idf (inverse document frequency) [10] によるランキングである。idf は、キーワードの重要性を扱う上で有効な頻度統計量であり、BM25 [9] など、ランキング関数などにも広く応用されている。本実験では、対象としている全映像集合における、キーワードが付与された映像数の逆数が大きいものから、降順にランキングするものとした。

RR : 適合性のみによるランキングである。3.1 節で示した、キーワードと映像内容との適合性の式(4)の値の降順にランキングするものとした。

CR : 具体性のみによるランキングである。3.2 節で示した、キーワードの具体性の式(6)の値の降順にランキングするものとした。

4.3 YouTube データによる実験

YouTube の映像データセット 50 本における、ランキングとタグの適合性および具体性の正解データセットとの順位相関係数の平均値を図 2 に示す。タグの適合性、具体性の両指標について、提案手法、RR、CR ともにベースラインである idf と比べてよく、中でも提案手法が最もよい結果となった。また、RR と CR を比較すると、タグの適合性の項目では RR がよく、タグの具体性の項目では CR がよいという結果となった。ここで、4.1 節に示したように、評価者によるランキング結果同士の順位相関係数の平均は、適合性で約 0.556、具体性で 0.588 であった。提案手法の結果と比較すると、具体性は 0.564 と、評価者同士の場合と比べ、遜色のない結果となったものの、適合性は 0.388 と差がみられるため、適合性の算出方法については、更なる改善が必要であると考える。

4.4 IPTV データによる実験

IPTV の映像データセット 50 本における結果を、図 3 に示す。キーワードの適合性、具体性のどちらの指標についても、提案手法、RR、CR ともにベースラインと比べてよい結果となった。しかし、YouTube の実験とは異なり、提案手法、RR、CR の間に大きな差はなかった。この理由を以下に考察する。

正解データにおける、適合性と具体性のランキングの順位相関係数を求めたところ、YouTube の場合では約 0.665 であったのに対し、IPTV の場合では約 0.812 と高かった。すなわち、IPTV データにおいて、具体性の高いキーワードは、同時に適合性も高く、このため、RR、CR、および、提案手法のランキング結果に大きな差がなかったと考えられる。図 4 に IPTV の番組概要文の一例と、ここから抽出されたキーワードの CR によるランキング結果上位 5 件を示す。上位 5 件は、「太陽系、速度、途方、宇宙、解明」であった。これらのキーワードには具体的であると同時に、内容にも関係するキーワードが多く含まれている。一方で、YouTube データの場合には、IPTV データの場合と比べ、適合性と具体性の間に高い相関はなく、このため、両手法を組み合わせた場合に、タグの適合性および具体性のどちらにおいても、最もよい結果になったと考えられる。

5. まとめ

本論文では、映像に付与されたキーワードが映像内容との適合性および具体性の双方が高い順にキーワードをランキングする手法を提案した。本手法は、映像に付与された

ザ・宇宙 ～神秘と驚異～「宇宙のスピード」

放送局 ナショナル ジオグラフィック チャンネル

使える土地を広げるには、移動手段のスピードを上げることが必要だ。太陽系の果てまで行ってみるにしても、大きさと時間の謎を解明するにしても、カギとなるのはスピードだ。単に素早い動きから想像を絶する速さまでスピードの世界を探し、途方もない速度によってとれた宇宙の謎が解明され、想像を超えた世界を見ることができると検証する

CRによるキーワードのランキング結果 上位5件

太陽系 速度 途方 宇宙 解明

図 4 IPTV の番組概要文の例

キーワード間の共起性の分析により、内容との適合性を推定し、また、共起性のばらつき分析により、具体性を推定する。これらの 2 つの推定結果を組み合わせることににより、内容に適合し、かつ具体性の高いキーワードから順にランキングする。約 300,000 本の IPTV の番組情報と約 50,000 本の YouTube の映像情報を用いて、評価実験を実施した。結果、提案手法が、従来の idf によるキーワードランキングに比べ、適合性、具体性の双方の観点でよいキーワードを上位にランキングできることを確認した。

今回の実験では、適合性と具体性の組み合わせ方式(1)において、それぞれの重みを均等にしておき、式(6)のパラメータ σ も最適化を行っていない。パラメータを変化させた場合における精度の変化の調査は今後の課題である。また、提案手法を利用することにより検索精度がどの程度向上するかについて、今後検証していく予定である。

謝辞

本研究を遂行するにあたり、有益なご示唆・ご協力をいただいた、NTT サイバーソリューション研究所 加藤洋一氏、および、中山彰氏に感謝いたします。

参考文献

- [1] D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang. Tag ranking. In Proc. WWW, pp. 351–360, 2009.
- [2] T. Iwata, T. Yamada, and N. Ueda. Modeling Social Annotation Data with Content Relevance using a Topic Model. In Proc. NIPS, pp. 835–843, 2009.
- [3] L. Wu, L. Yang, N. Yu, and X.-S. Hua. Learning to tag. In Proc. WWW, pp. 361–370, 2009.
- [4] E. Moxley, T. Mei, X.-S. Hua, W.-Y. Ma, and B. Manjunath. Automatic Video Annotation Through Search and Mining. In Proc. ICME, pp. 685–688, 2008.
- [5] S. Siersdorfer, J.S. Pedro, and M. Sanderson. Automatic Video Tagging using Content Redundancy. In Proc. SIGIR, pp. 395–402, 2009.
- [6] B. Sigurbjörnsson and R. van Zwol. Flickr Tag Recommendation based on Collective Knowledge. In Proc. WWW, pp. 327–336, 2008.
- [7] Z. Chen, J. Cao, Y. Song, J. Guo, Y. Zhang, and J. Li. Context-Oriented Web Video Tag Recommendation. In Proc. WWW, pp. 1079–1080, 2010.
- [8] G. Salton and C. Buckley. Term-weighting Approaches in Automatic Text Retrieval. Information Processing and Management, Vol.24, No.5, pp. 513–523, 1988.
- [9] S.E. Robertson, S. Walker, S. Jones, M.M. Hancock-Beaulieu, and M. Gattford. Okapi at TREC-3. In Proc. TREC-3, pp. 109–126, 1994.