

コーパス検索と著作権

田淵 龍二 (ミント音声教育研究所)



発表要項

http://mintap.kir.jp/public/news/pic/let_201808.pdf



ハンドアウト

キーワード： コーパス， 検索エンジン， 著作権， 言語資源， 公開

1. はじめに

コーパスを使った研究や学習の機会が増大しているが、円滑にアクセスするための検索システムが必須で、Google などの検索サイトが有名である。検索エンジンの開発は、科研費研究の対象分野でもある。

他方、コーパスは著作物なので著作権の問題が生起するが、法律が IT に対応しきれずおらず混乱や萎縮があり、「教育機関の理解が不十分」（文化庁, 2017）とされる。



https://www.kantei.go.jp/jp/singi/tiiki/kokusentoc_wg/h28/shouchou/170118_shiryoushiryo_us41.pdf

文化庁 (2017) p.4

2. 目的

コーパス構築とコーパス検索エンジン開発における著作権の処理実態をあきらかにする。

3. 方法

コーパス構築の専門家が集まる言語処理学会の論文 4,428 本から、文中に「コーパス」と「著作権」を含む論文を、論文検索サイト NaCSE で検索し、精査した。



<http://www.mintap.com/nacse/nacse.html>

言語処理学会論文コーパス **NaCSE 3**

4. 結果

コーパスと著作権を同時に含む論文は 66 本あり，本文に共起した論文は 40 本であった。

これらを調査したところ，構築したコーパスにおける著作権の扱いは 7 つに分類された (表 1)。

また，著作権処理の扱いを決めた要因を調べたところ 7 つの因子が明らかとなった (表 2)。

コーパス構築における著作権処理の仕方と決定要因

表 1.

コーパス構築における著作権の扱い

1. 著作権処理を実施
 2. 著作権保護期間終了の著作物
 3. コーパス未公開
 4. 解析結果 (例: n-gram) のみ公開
 5. 元著作物へのリンクを公開
 6. (疑似) テキストを生成して公開
 7. 公開可能な著作物
-

表 2.

著作権の扱いを左右した主な要因

- A. 著作物が自前か否か
 - B. 著作権保護期間内か否か
 - C. 著作権者が少数か膨大か
 - D. 著作物がウェブ公開サイトか否か
 - E. 著作物を公開するか否か
 - F. 必要なのは著作物本体か解析結果か
 - G. 製作者の予算と人員の大小
-

5. 考察

コーパスと検索エンジンを公開するか否かは開発目的により異なり，公開可能か否かは個々の著作物により異なる。

しかし，著作権問題に萎縮して放棄あるいは禁止したり，さらには法律論に深入りするよりも，表 2 で見た事情に応じて，表 1 の対処方法を目安にすることが現実的であり，著作権法の精神「この法律は，・・・著作者等の権利の保護を図り，もつて文化の発展に寄与することを目的とする」（第一条）に沿うものではないかと思量される。

フェアユースの必然性

5 年の歳月をかけ、2 万件を越える著作権処理を行った前川(国立国語研究所)は「フェアユース条項の導入は長期的には必然」としている。

しかし筆者はフェアユース条項の導入は長期的には必然であろうと考えている。現在の著作権法は、著作物の権利者が社会全体からみれば少数のプロフェッショナルに限られている時代に作られたものであるため、インターネットに代表される安価な情報発信手段の爆発的普及によって国民の多くが著作権者と化した情報化社会（それは仮想社会の話ではなく、今私たちの目の前に広がっている現実である）においては、もはや個別的な制限規定では急速に変化（あえて進化とは言わない）しつづける現実に対応することができないからである。

コーパス構築と著作権保護 (前川 2010)



http://pj.ninjal.ac.jp/corpus_center/bccwj/doc/Maekawa2010.pdf

フェアユース (fair use)

「公正な利用であれば、著作権者の許諾なく著作物を利用できる」とする法律規定

- 米国の場合は 著作権法 107 条
- 日本では、これまで「個別具体的な制限規定」（例えば 35 条学校での複製等）であったが、今年 5 月に法改正があり「柔軟な権利制限規定」が導入された。

YouTube とフェアユース

YouTube に投稿された動画をフェアユースの観点から保護する指針を公開

YouTube は、クリエイティブな動画環境全体をより良いものにすることを目的として、裁判所の指示により、著作権者が削除通知を送信する前にフェアユースに該当するかどうかの分析を行ったかどうかを確認している。

(YouTube / フェアユース)



<https://www.youtube.com/intl/ja/yt/about/copyright/fair-use/#yt-copyright-resources>

クリエイティブ・コモンズ・ライセンス (CCL)

Creative Commons License

著作権者が、著作権を保持したまま、作品の使用を他者に許可するための方法。

YouTube では、投稿した動画に CCL (CC BY) を適用することで、ユーザーに動画を再利用 (編集も含む) してもらうことを目的としている。
(YouTube ヘルプ / クリエイティブ・コモンズ)



<https://support.google.com/youtube/answer/2797468?hl=ja>

TED とクリエイティブ・コモンズ・ライセンス

TED は TED Talks にクリエイティブ・コモンズ (CC BY-NC-ND ; 表示・非営利・改変禁止) を適用している。

(TED Talks Usage Policy / TED Talks under Creative Commons License)



<https://www.ted.com/about/our-organization/our-policies/terms/ted-talks-usage-policy>

TEDxTokyo も自らのビデオに CCL を適用している

「誰でも自由に TED.com と tedxtokyo.com からビデオをダウンロードすることができます。友達と共有したり、ウェブサイトやブログに埋め込んで自由に使ってください。」(TEDxTokyo / クリエイティブ・コモンズ)



<https://www.tedxtokyo.com/creative-commons/?lang=ja>

国が検討中の著作権法改正 (2018 年 5 月に成立)

デジタル化・ネットワーク化の進展に対応した 柔軟な権利制限規定の整備

著作権法の一部を改正する法律案 概要説明資料 (A I の利活用促進関係)

情報通信技術の進展等の著作物等の利用をめぐる環境の変化に対応し、著作物等の公正な利用を図るとともに著作権等の適切な保護に資するため、電子計算機における著作物の利用に付随する利用、学校その他の教育機関における公衆送信、美術の著作物等の展示に伴う複製等をより円滑に行えるようにするための措置等を講ずるほか、盲人、視覚障害者その他の印刷物の判読に障害のある者が発行された著作物を利用する機会を促進するためのマラケシュ条約(仮称)に対応するため、視覚障害者等に係る権利制限規定の対象者の範囲を拡大する。

- I. デジタル化・ネットワーク化の進展に対応した柔軟な権利制限規定の整備
【30条の4、47条の4、47条の5等関係】
- II. 教育の情報化に対応した権利制限規定等の整備【35条等関係】 (略)
- III. 障害者の情報アクセス機会の充実に係る権利制限規定の整備【37条関係】 (略)
- IV. アーカイブの利活用促進に関する権利制限規定の整備等【第31条、第47条、第67条等関係】 (略)

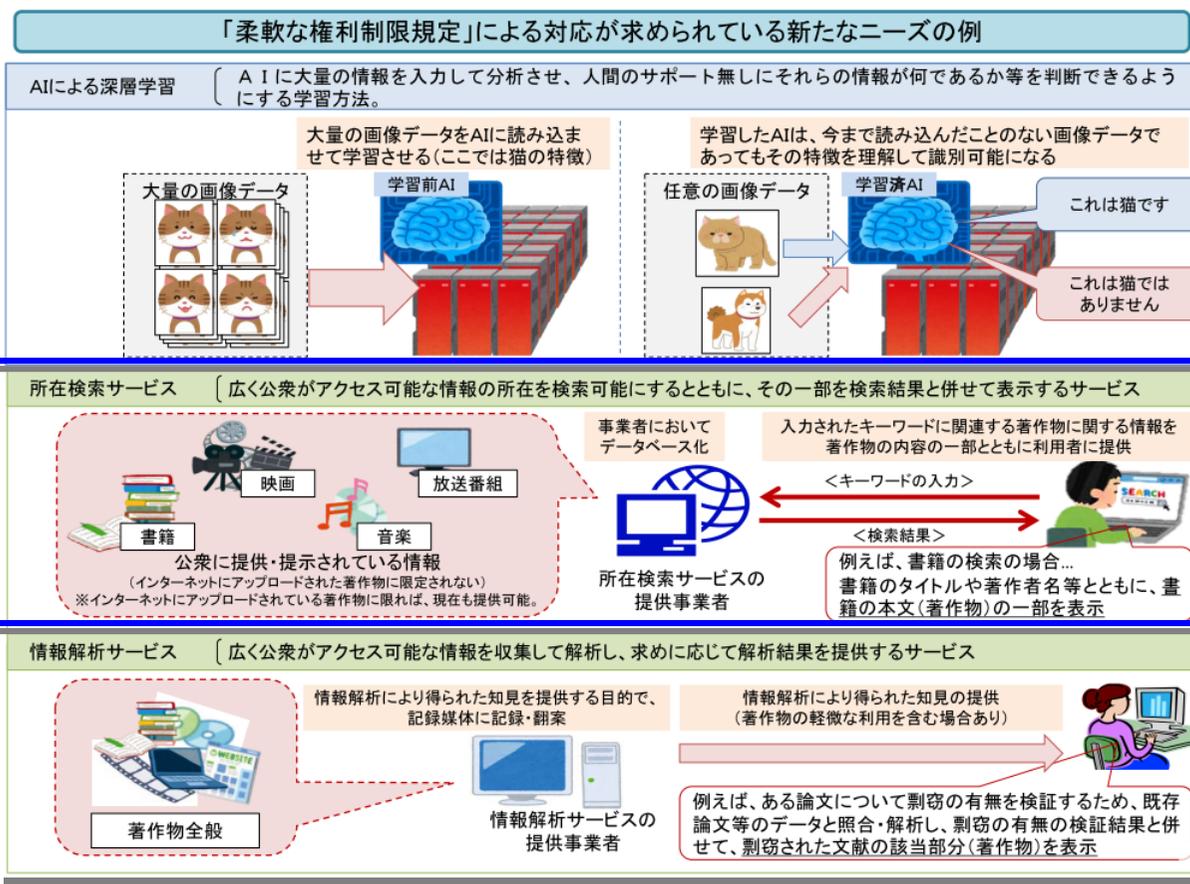
平成30年4月2日
文化庁長官官房著作権課



https://www.kantei.go.jp/jp/singi/titeki2/tyou_sakai/kensho_hyoka_kikaku/2018/sangyou/daishu/siryou2-4.pdf

2018 年 4 月 2 日 文化庁長官官房著作権課

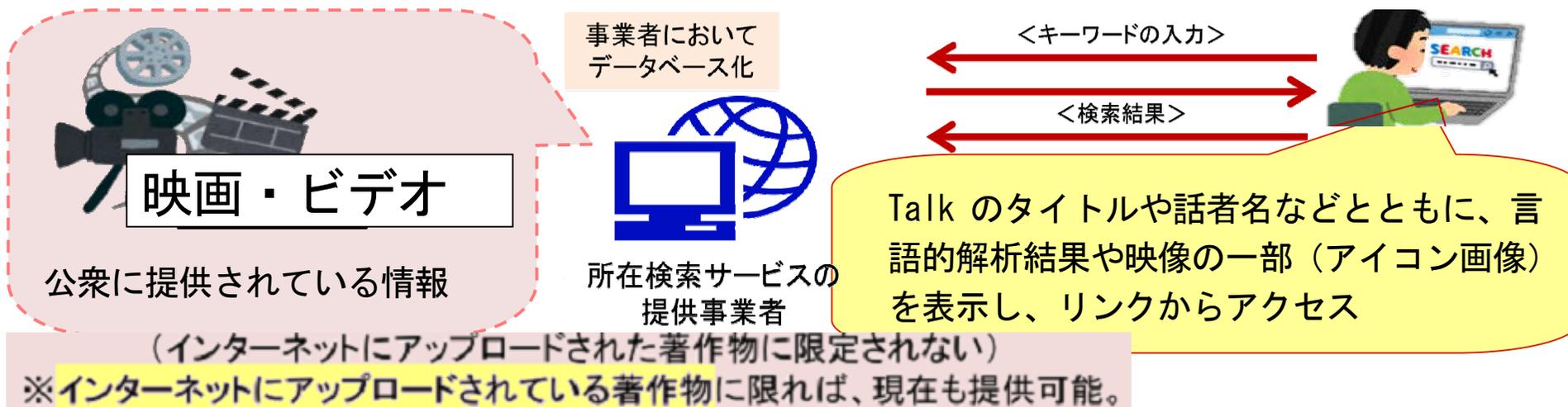
所在検索サービス>アップロード著作物に限定しない 新たなニーズ



2018年4月2日 文化庁長官官房著作権課

所在検索サービスは現法律ですでに提供可能となっている

TED Talks Search Engine “selected360” の場合



検索エンジンを明記した 2009 年著作権法改正

著作権法改正概要(文化庁 2009)

1. インターネット等を活用した著作物利用の円滑化を図るための措置

(1) インターネット情報検索サービスを実施するための複製等 (47条の9)

インターネットでの情報検索サービス (Yahoo!やGoogle) に伴う、情報の収集、整理・解析、検索結果の表示が、著作権法に抵触する可能性があり、日本国内にサーバーを設置できないとの指摘。 (「情報大航海プロジェクト (経済産業省)」を推進する観点からも、検討の要請。)

法改正の内容

情報検索サービスに必要な行為は、著作権者の許諾を得なくても可能とすることを明確化する。ただし、次の条件を付する。

①権利者がネット上で情報収集を拒否する旨の意思表示を行っている場合は、当該情報を収集しない。(※政令で規定)

②サービス事業者が違法複製物の存在を知った場合、その表示を停止する。

国内でも安心して情報検索サービスが実施できるようになり、次世代サービス開発が加速。



http://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h21_hokaisei/pdf/21_hokaisei_hori_tsu_gaiyou.pdf

著作権侵害で訴えられる恐れがあるから教育研究活動をしな
い・・・との LET の態度は、もはや時代に合わない。

1. 検索サイト＝適法（日本 2010 年以降）なので侵害しない
2. 半年以上やり取りしている・・・訴えられていない
3. 著作権への理解不足と萎縮（文科省）

ご清聴ありがとうございました



ミント音声教育研究所 田淵龍二
Mail: tabuchiryuji@nifty.ne.jp