

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5051237号  
(P5051237)

(45) 発行日 平成24年10月17日(2012.10.17)

(24) 登録日 平成24年8月3日(2012.8.3)

(51) Int.Cl. F I  
**G06F 21/22 (2006.01)** G O 6 F 21/22 1 1 O L  
**G06F 17/30 (2006.01)** G O 6 F 17/30 1 2 O A  
 G O 6 F 17/30 3 5 O C

請求項の数 50 (全 28 頁)

<p>(21) 出願番号 特願2009-537912 (P2009-537912)                  (86) (22) 出願日 平成20年10月14日 (2008.10.14)                  (86) 国際出願番号 PCT/JP2008/002899                  (87) 国際公開番号 W02009/050877                  (87) 国際公開日 平成21年4月23日 (2009.4.23)                  審査請求日 平成23年9月6日 (2011.9.6)                  (31) 優先権主張番号 特願2007-272968 (P2007-272968)                  (32) 優先日 平成19年10月19日 (2007.10.19)                  (33) 優先権主張国 日本国(JP)</p>	<p>(73) 特許権者 000004237                  日本電気株式会社                  東京都港区芝五丁目7番1号                  (74) 代理人 100110928                  弁理士 速水 進治                  (72) 発明者 平田 恭二                  東京都港区芝五丁目7番1号 日本電気株式会社社内                  (72) 発明者 芹沢 昌宏                  東京都港区芝五丁目7番1号 日本電気株式会社社内                  審査官 岸野 徹</p>
---	--

最終頁に続く

(54) 【発明の名称】 不適切コンテンツ検出方法及び装置、そのコンピュータプログラム、ならびにコンテンツ公開システム

(57) 【特許請求の範囲】

【請求項1】

不適切コンテンツ検出装置が、  
 個々の利用者から投稿されたコンテンツを受け付け、  
 受け付けた複数の前記投稿されたコンテンツを用いて、複数の前記投稿されたコンテンツの相互の類似度を算出し、  
 前記相互の類似度に基づき、前記投稿されたコンテンツが著作権的に適切であるか否かの判定を行う不適切コンテンツ検出方法。

【請求項2】

不適切コンテンツ検出装置が、  
 個々の利用者から投稿されたコンテンツを受け付け、  
 受け付けた複数の前記投稿されたコンテンツを用いて、複数の前記投稿されたコンテンツ間で算出された相互の類似度に基づき、複数の前記投稿されたコンテンツが相互に類似しているか否かを判定し、類似していた場合に、相互に類似しているコンテンツ群を著作権的に不適切なコンテンツとして検出する不適切コンテンツ検出方法。

【請求項3】

請求項1または2に記載の不適切コンテンツ検出方法において、  
 前記不適切コンテンツ検出装置が、  
 所定の期間内に投稿されたコンテンツのうち、所定数以上の前記投稿されたコンテンツ間で相互に前記類似度が閾値より高いと判定されたコンテンツ群を、前記不適切なコンテ

ンツとして検出する不適切コンテンツ検出方法。

【請求項 4】

請求項 1 乃至 3 いずれか一項に記載の不適切コンテンツ検出方法において、前記不適切コンテンツ検出装置が、

複数の前記投稿されたコンテンツから、所定の基準でコンテンツを選択した後に、前記選択されたコンテンツ間で相互に類似度を算出し、所定数以上のコンテンツ間で、相互に類似度が閾値より高いと判定されたコンテンツ群を、前記不適切なコンテンツとして検出する不適切コンテンツ検出方法。

【請求項 5】

請求項 1 乃至 4 いずれか一項に記載の不適切コンテンツ検出方法において、前記不適切コンテンツ検出装置が、

前記投稿されたコンテンツから特徴量を抽出し、前記投稿されたコンテンツの前記特徴量を、相互に照合することでコンテンツ相互の類似度を算出する不適切コンテンツ検出方法。

【請求項 6】

請求項 1 乃至 5 いずれか一項に記載の不適切コンテンツ検出方法において、前記不適切コンテンツ検出装置が、

前記投稿されたコンテンツが映像であった際に、映像に含まれる一部または全部のフレーム群が相互に類似しているときに、コンテンツ間の類似度が閾値より高いとする不適切コンテンツ検出方法。

【請求項 7】

請求項 1 乃至 6 いずれか一項に記載の不適切コンテンツ検出方法において、前記不適切コンテンツ検出装置が、

前記投稿されたコンテンツが静止画像であった際に、静止画像の一部または全体が類似しているときに、コンテンツ間の類似度が閾値より高いとする不適切コンテンツ検出方法。

【請求項 8】

請求項 1 乃至 7 いずれか一項に記載の不適切コンテンツ検出方法において、前記不適切コンテンツ検出装置が、

前記投稿されたコンテンツが音響または音楽を含む際に、音響または音楽の一部または全体のフレーズが類似しているときに、コンテンツ間の類似度が閾値より高いとする不適切コンテンツ検出方法。

【請求項 9】

請求項 1 乃至 8 いずれか一項に記載の不適切コンテンツ検出方法において、前記不適切コンテンツ検出装置が、特徴量蓄積部を備え、前記不適切コンテンツ検出装置が、

投稿毎に前記投稿されたコンテンツの特徴量を抽出して前記特徴量蓄積部に蓄積しておき、新たに投稿された新規コンテンツと複数の前記蓄積されたコンテンツの前記特徴量に基づいて、複数の前記蓄積されたコンテンツ毎に、前記新規コンテンツとの類似度を算出し、前記新規コンテンツとの前記類似度が閾値より高いコンテンツが存在した場合に、前記新規コンテンツおよび、前記新規コンテンツとの前記類似度が閾値より高いと判定されたコンテンツのコンテンツ群を前記不適切なコンテンツと判定する不適切コンテンツ検出方法。

【請求項 10】

請求項 1 乃至 8 いずれか一項に記載の不適切コンテンツ検出方法において、前記不適切コンテンツ検出装置が、特徴量蓄積部を備え、前記不適切コンテンツ検出装置が、

これまでに投稿されたコンテンツの特徴量を抽出して前記特徴量蓄積部に蓄積しておき、新たに投稿された新規コンテンツと複数の前記蓄積されたコンテンツの前記特徴量に基づいて、複数の前記蓄積されたコンテンツ毎に、前記新規コンテンツとの類似度を算出し

10

20

30

40

50

、前記新規コンテンツとの前記類似度が閾値より高いコンテンツが存在した場合に、前記蓄積されたコンテンツに前記新規コンテンツを加えて、相互の前記類似度を算出し、所定数以上のコンテンツ間で相互に前記類似度が閾値より高いと判定されたコンテンツ群を、前記不適切なコンテンツとして検出する不適切コンテンツ検出方法。

【請求項 1 1】

請求項 1 0 に記載の不適切コンテンツ検出方法において、  
前記不適切コンテンツ検出装置が、

前記新規コンテンツと比較する前記蓄積されたコンテンツとして、所定期間に投稿されたコンテンツのみを用いる不適切コンテンツ検出方法。

【請求項 1 2】

請求項 1 乃至 1 1 いずれか一項に記載の不適切コンテンツ検出方法において、  
前記不適切コンテンツ検出装置が、辞書記憶部を備え、  
前記不適切コンテンツ検出装置が、

前記不適切と判定された投稿コンテンツまたは該コンテンツの特徴量を不適切コンテンツ辞書データとして前記辞書記憶部に記憶し、新規に投稿されたコンテンツに対して、前記記憶された不適切コンテンツ辞書データと照合を行うことにより前記不適切なコンテンツの検出を行う不適切コンテンツ検出方法。

【請求項 1 3】

請求項 1 2 に記載の不適切コンテンツ検出方法において、  
前記不適切コンテンツ検出装置が、

所定期間または利用者またはコンテンツに対する相互の類似度の算出により前記不適切コンテンツ辞書データを生成する不適切コンテンツ検出方法。

【請求項 1 4】

請求項 1 2 または 1 3 に記載の不適切コンテンツ検出方法において、  
前記不適切コンテンツ検出装置が、

前記相互の類似度算出による不適切コンテンツ検出を定期的に行い前記不適切コンテンツ辞書データの生成または更新を行う不適切コンテンツ検出方法。

【請求項 1 5】

請求項 1 2 乃至 1 4 いずれか一項に記載の不適切コンテンツ検出方法において、  
前記不適切コンテンツ検出装置が、

前記不適切コンテンツ辞書データを情報管理者が確認し、前記情報管理者が指示した不適切コンテンツのみを保持するよう修正する不適切コンテンツ検出方法。

【請求項 1 6】

請求項 1 乃至 1 5 いずれか一項に記載の不適切コンテンツ検出方法において、  
前記不適切コンテンツ検出装置が、登録部を備え、  
前記不適切コンテンツ検出装置が、

所定の利用者または所定のコンテンツを前記登録部に予め登録しておき、前記所定の利用者により投稿されたコンテンツまたは前記所定のコンテンツは前記不適切コンテンツとして検出しない不適切コンテンツ検出方法。

【請求項 1 7】

個々の利用者から投稿されたコンテンツの入力を受け付けるコンテンツ受付手段と、受け付けた複数の前記投稿されたコンテンツを用いて、複数の前記投稿されたコンテンツの相互の類似度を算出し、前記類似度に基づき、相互に類似するコンテンツ群を検出する類似コンテンツ検出手段と、

検出された前記類似するコンテンツ群に基づき、著作権的に不適切なコンテンツを判定する不正判定手段と、を備える不適切コンテンツ検出装置。

【請求項 1 8】

請求項 1 7 に記載の不適切コンテンツ検出装置において、

前記類似コンテンツ検出手段は、所定の期間内に投稿されたコンテンツのうち、所定数以上の前記投稿されたコンテンツ間で相互に前記類似度が閾値より高いコンテンツを前記

10

20

30

40

50

相互に類似するコンテンツ群として検出し、

前記不正判定手段は、検出された前記類似するコンテンツ群を前記不適切なコンテンツと判定する不適切コンテンツ検出装置。

【請求項 19】

請求項 17 または 18 に記載の不適切コンテンツ検出装置において、

前記類似コンテンツ検出手段は、複数の前記投稿されたコンテンツから、所定の基準でコンテンツを選択した後に、前記選択されたコンテンツ間で相互に類似度を算出し、所定数以上のコンテンツ間で、相互に類似度が閾値より高いコンテンツ群を検出し、

前記不正判定手段は、検出された前記コンテンツ群を、前記不適切なコンテンツと判定する不適切コンテンツ検出装置。

10

【請求項 20】

請求項 17 乃至 19 いずれか一項に記載の不適切コンテンツ検出装置において、

前記投稿されたコンテンツからその特徴量を抽出する特徴量抽出手段を備え、

前記類似コンテンツ検出手段は、複数の前記投稿されたコンテンツの前記特徴量を相互に照合して、前記特徴量の相互の類似度を算出し、前記類似度に基づいて、相互に類似するコンテンツ群を検出する不適切コンテンツ検出装置。

【請求項 21】

請求項 20 に記載の不適切コンテンツ検出装置において、

前記特徴量抽出手段によって抽出された前記特徴量を記憶する特徴量蓄積手段と、

前記特徴量蓄積手段から、所定の基準で選択されたコンテンツの特徴量を取得する蓄積特徴量選択手段と、を備え、

前記類似コンテンツ検出手段は、前記蓄積特徴量選択手段によって選択されたコンテンツの前記特徴量の相互の類似度を算出し、相互に類似するコンテンツを検出する不適切コンテンツ検出装置。

20

【請求項 22】

請求項 17 乃至 21 いずれか一項に記載の不適切コンテンツ検出装置において、

前記類似コンテンツ検出手段は、前記投稿されたコンテンツが映像であった際に、映像に含まれる一部または全部のフレーム群が相互に類似しているときに、コンテンツ間の類似度が閾値より高いとする不適切コンテンツ検出装置。

【請求項 23】

請求項 17 乃至 22 いずれか一項に記載の不適切コンテンツ検出装置において、

前記類似コンテンツ検出手段は、前記投稿されたコンテンツが静止画像であった際に、静止画像の一部または全体が類似しているときに、コンテンツ間の類似度が閾値より高いとする不適切コンテンツ検出装置。

30

【請求項 24】

請求項 17 乃至 23 いずれか一項に記載の不適切コンテンツ検出装置において、

前記類似コンテンツ検出手段は、前記投稿されたコンテンツが音響または音楽を含む際に、音響または音楽の一部または全体のフレーズが類似しているときに、コンテンツ間の類似度が閾値より高いとする不適切コンテンツ検出装置。

【請求項 25】

請求項 21 に記載の不適切コンテンツ検出装置において、

前記特徴量蓄積手段は、投稿毎に前記投稿されたコンテンツの前記特徴量を蓄積しておく、

前記類似コンテンツ検出手段は、新たに投稿された新規コンテンツと複数の前記蓄積されたコンテンツの前記特徴量に基づいて、複数の前記蓄積されたコンテンツ毎に、前記新規コンテンツとの類似度を算出し、前記新規コンテンツとの前記類似度が閾値より高いコンテンツが存在した場合に、前記新規コンテンツおよび、前記新規コンテンツとの前記類似度が閾値より高いコンテンツのコンテンツ群を検出し、

40

前記不正判定手段は、検出された前記コンテンツ群を前記不適切なコンテンツと判定する不適切コンテンツ検出装置。

50

## 【請求項 2 6】

請求項 2 1 に記載の不適切コンテンツ検出装置において、

前記特徴量蓄積手段は、これまでに投稿されたコンテンツの前記特徴量を蓄積しておき

、  
前記類似コンテンツ検出手段は、新たに投稿された新規コンテンツと複数の前記蓄積されたコンテンツの前記特徴量に基づいて、複数の前記蓄積されたコンテンツ毎に、前記新規コンテンツとの類似度を算出し、前記新規コンテンツとの前記類似度が閾値より高いコンテンツが存在した場合に、前記蓄積されたコンテンツに前記新規コンテンツを加えて、相互の前記類似度を算出し、所定数以上のコンテンツ間で相互に前記類似度が閾値より高いコンテンツ群を検出し、

10

前記不正判定手段は、検出された前記コンテンツ群を前記不適切なコンテンツと判定する不適切コンテンツ検出装置。

## 【請求項 2 7】

請求項 2 6 に記載の不適切コンテンツ検出装置において、

前記類似コンテンツ検出手段は、前記新規コンテンツと比較する前記蓄積されたコンテンツとして、所定期間に投稿されたコンテンツのみを用いる不適切コンテンツ検出装置。

## 【請求項 2 8】

請求項 1 7 乃至 2 7 いずれか一項に記載の不適切コンテンツ検出装置において、

前記不正判定手段で不適切と判定された前記不適切なコンテンツまたは該不適切なコンテンツの特徴量を不適切コンテンツ辞書データとして記憶する辞書記憶手段と、

20

新規に投稿されたコンテンツに対して、前記不適切コンテンツ辞書データと照合を行うことにより前記不適切なコンテンツの検出を行う不適切コンテンツ検出手段と、をさらに備える不適切コンテンツ検出装置。

## 【請求項 2 9】

請求項 2 8 に記載の不適切コンテンツ検出装置において、

所定期間または所定の利用者または所定のコンテンツに対する相互の類似度の算出により前記不適切コンテンツ辞書データを生成する生成手段をさらに備える不適切コンテンツ検出装置。

## 【請求項 3 0】

請求項 2 9 に記載の不適切コンテンツ検出装置において、

前記生成手段は、前記相互の類似度の算出による不適切コンテンツ検出を定期的に行い、前記不適切コンテンツ辞書データの生成または更新を行う不適切コンテンツ検出装置。

30

## 【請求項 3 1】

請求項 2 8 乃至 3 0 いずれか一項に記載の不適切コンテンツ検出装置において、

前記不適切コンテンツ辞書データを情報管理者に提示し、該情報管理者が前記不適切コンテンツ辞書データを確認し、保持する不適切コンテンツ辞書データの指示を受け付け、指示された不適切コンテンツ辞書データのみを前記辞書記憶手段に記憶する不適切コンテンツ検出装置。

## 【請求項 3 2】

請求項 1 7 乃至 3 1 いずれか一項に記載の不適切コンテンツ検出装置において、

所定の利用者または所定のコンテンツを予め登録する登録手段を備え、

前記不正判定手段は、前記所定の利用者により投稿されたコンテンツまたは前記所定のコンテンツは前記不適切なコンテンツと判定しない不適切コンテンツ検出装置。

40

## 【請求項 3 3】

コンピュータに、個々の利用者から投稿されたコンテンツから著作権的に不適切なコンテンツを検出する不適切コンテンツ検出装置を実現させるためのコンピュータプログラムであって、

前記コンピュータに、

個々の利用者から投稿されたコンテンツの入力を受け付ける手順と、

受け付けた複数の前記投稿されたコンテンツを用いて、複数の前記投稿されたコンテン

50

ツの相互の類似度を算出し、前記類似度に基づき、相互に類似するコンテンツ群を検出する手順、

検出された前記類似するコンテンツ群に基づき、著作権的に不適切なコンテンツを判定する手順、を実行させるためのコンピュータプログラム。

【請求項 3 4】

請求項 3 3 に記載のコンピュータプログラムにおいて、

前記コンピュータに、

所定の期間内に投稿されたコンテンツのうち、所定数以上の前記投稿されたコンテンツ間で相互に前記類似度が閾値より高いコンテンツを前記相互に類似するコンテンツ群として検出する手順をさらに実行させるためのコンピュータプログラム。

10

【請求項 3 5】

請求項 3 3 または 3 4 に記載のコンピュータプログラムにおいて、

前記コンピュータに、

前記投稿されたコンテンツからその特徴量を抽出する手順、

複数の前記投稿されたコンテンツの前記特徴量を相互に照合して、前記特徴量の相互の類似度を算出する手順、

算出された前記類似度に基づいて、前記相互に類似するコンテンツ群を検出する手順、をさらに実行させるためのコンピュータプログラム。

【請求項 3 6】

請求項 3 3 乃至 3 5 いずれか一項に記載のコンピュータプログラムにおいて、

20

前記コンピュータは、不適切コンテンツ辞書データを記憶する辞書記憶装置を備え、

前記コンピュータに、

前記不適切なコンテンツを判定する手順で不適切と判定された前記不適切なコンテンツまたは該不適切なコンテンツの特徴量を前記不適切コンテンツ辞書データとして前記辞書記憶装置に記憶する手順、

新規に投稿されたコンテンツに対して、前記不適切コンテンツ辞書データと照合を行うことにより前記不適切なコンテンツの検出を行う手順、をさらに実行させるためのコンピュータプログラム。

【請求項 3 7】

投稿されたコンテンツを利用者が閲覧できるように公開するコンテンツ公開システムであって、

30

請求項 1 7 乃至 3 2 いずれか一項に記載の不適切コンテンツ検出装置によって検出された前記不適切なコンテンツをシステム管理者に提示する提示手段と、

前記システム管理者が確認した後、前記システム管理者から削除指示を受け付ける受付手段と、

前記削除指示に従い、前記不適切なコンテンツの削除を行う削除手段と、を備えるコンテンツ公開システム。

【請求項 3 8】

投稿されたコンテンツを利用者が閲覧できるように公開するコンテンツ公開システムであって、

40

請求項 1 7 乃至 3 2 いずれか一項に記載の不適切コンテンツ検出装置によって相互に類似していると判定されたコンテンツ数が、所定数より多いか否かを判定する判定手段と、

相互に類似していると判定されたコンテンツ数が所定数より多い場合に、当該コンテンツへの利用者のアクセスを自動的に停止する制御手段と、を備えるコンテンツ公開システム。

【請求項 3 9】

請求項 1 乃至 1 6 いずれか一項に記載の不適切コンテンツ検出方法において、

前記不適切コンテンツ検出装置が、

前記相互の類似度を算出するコンテンツは、著作権的に適切であるか否かが未確認の複数の投稿されたコンテンツである不適切コンテンツ検出方法。

50

## 【請求項 4 0】

請求項 1 2 乃至 1 5 いずれか一項に記載の不適切コンテンツ検出方法において、  
前記不適切コンテンツ検出装置が、  
前記不適切コンテンツ辞書データとして、著作権コンテンツを前記辞書記憶部に記憶する不適切コンテンツ検出方法。

## 【請求項 4 1】

請求項 1 乃至 1 6、3 9、および 4 0 いずれか一項に記載の不適切コンテンツ検出方法において、

前記不適切コンテンツ検出装置が、  
受け付けた複数の前記投稿されたコンテンツの中から順次コンテンツ対を選択し、  
選択された前記コンテンツ対の類似度を算出し、  
算出された前記類似度が予め定められた閾値以上のコンテンツの件数を求め、  
前記コンテンツの前記件数に基づいて、前記コンテンツが著作権的に適切であるか否かの判定を行う不適切コンテンツ検出方法。

10

## 【請求項 4 2】

請求項 4 1 に記載の不適切コンテンツ検出方法において、  
前記不適切コンテンツ検出装置が、  
算出された前記類似度が予め定められた閾値以上のコンテンツが、予め定められた件数以上検出された場合、検出されたコンテンツ対を類似コンテンツとして検出し、それらの類似コンテンツ群に対して、著作権的に不適切なコンテンツと判定する不適切コンテンツ  
検出方法。

20

## 【請求項 4 3】

請求項 1 7 乃至 3 2 いずれか一項に記載の不適切コンテンツ検出装置において、  
前記類似コンテンツ検出手段が前記相互の類似度を算出するコンテンツは、著作権的に適切であるか否かが未確認の複数の投稿されたコンテンツである不適切コンテンツ検出装置。

## 【請求項 4 4】

請求項 2 8 乃至 3 1 いずれか一項に記載の不適切コンテンツ検出装置において、  
前記辞書記憶手段は、前記不適切コンテンツ辞書データとして、著作権コンテンツを記憶する不適切コンテンツ検出装置。

30

## 【請求項 4 5】

請求項 1 7 乃至 3 2、4 3、および 4 4 いずれか一項に記載の不適切コンテンツ検出装置において、

前記類似コンテンツ検出手段は、  
前記受け付けた複数の前記投稿されたコンテンツの中から順次コンテンツ対を選択し、  
選択された前記コンテンツ対の類似度を算出し、  
算出された前記類似度が予め定められた閾値以上のコンテンツの件数を求め、  
前記不正判定手段は、  
前記コンテンツの前記件数に基づいて、前記コンテンツが著作権的に適切であるか否かの判定を行う不適切コンテンツ検出装置。

40

## 【請求項 4 6】

請求項 4 5 に記載の不適切コンテンツ検出装置において、  
前記類似コンテンツ検出手段は、算出された前記類似度が予め定められた閾値以上のコンテンツが、予め定められた件数以上検出された場合、検出されたコンテンツ対を類似コンテンツとして検出し、  
前記不正判定手段は、それらの類似コンテンツ群に対して、著作権的に不適切なコンテンツと判定する不適切コンテンツ検出装置。

## 【請求項 4 7】

請求項 3 3 乃至 3 6 いずれか一項に記載のコンピュータプログラムにおいて、

50

前記相互に類似するコンテンツ群を検出する手順で、前記相互の類似度を算出するコンテンツは、著作権的に適切であるか否かが未確認の複数の投稿されたコンテンツであるコンピュータプログラム。

【請求項 48】

請求項 36 に記載のコンピュータプログラムにおいて、  
前記コンピュータに、

前記不適切コンテンツ辞書データとして、著作権コンテンツを前記辞書記憶装置に記憶する手順をさらに実行させるためのコンピュータプログラム。

【請求項 49】

請求項 33 乃至 36、47、および 48 いずれか一項に記載のコンピュータプログラムにおいて、

前記コンピュータに、

受け付けた複数の前記投稿されたコンテンツの中から順次コンテンツ対を選択する手順

、  
選択された前記コンテンツ対の類似度を算出する手順、  
算出された前記類似度が予め定められた閾値以上のコンテンツの件数を求める手順、  
前記コンテンツの前記件数に基づいて、前記コンテンツが著作権的に適切であるか否かの判定を行う手順をさらに実行させるためのコンピュータプログラム。

【請求項 50】

請求項 49 に記載のコンピュータプログラムにおいて、

前記コンピュータに、

算出された前記類似度が予め定められた閾値以上のコンテンツが、予め定められた件数以上検出された場合、検出されたコンテンツ対を類似コンテンツとして検出する手順、

それらの類似コンテンツ群に対して、著作権的に不適切なコンテンツと判定する手順をさらに実行させるためのコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、投稿サイトにおける不適切コンテンツ検出方法、不適切コンテンツ検出装置、そのコンピュータプログラム、およびコンテンツ公開システムに関する。

【背景技術】

【0002】

インターネットなどにて展開されるサービスの一つに、掲示板サービスあるいは動画または静止画投稿サービスがある。掲示板等のこれらのサービスでは、不特定多数の利用者が、画像、音声、テキスト等のデータ（コンテンツ）をアップロードしたり、他人のアップロードしたコンテンツを自由に閲覧したりすることができる。

【0003】

このような掲示板等への投稿を、投稿する利用者の自由に任せると、他人の著作権物をコピーまたは模造したいわゆる不正コンテンツや不法コンテンツが大量に出回ってしまう危険性がある。現在不正コンテンツは、サービス提供者が定期的に監視を行うか、または著作権者がサービス提供者にクレームして該コンテンツを削除するといった枠組みで運用されているが、掲載されるコンテンツが大量になった場合、そのすべてを手でチェックすることは困難である。

【0004】

投稿されたコンテンツの中から不適切なデータを検出するシステムの一例が特許文献 1 に記載されている。同文献に記載されたシステムは、予め不適切なコンテンツ（画像、音声、テキスト等）をみつけるためのベースとなるコンテンツのサンプルを蓄えておき、新規にコンテンツ（画像、音声、テキスト等）が投稿されるたびに、前記蓄えられた不適切なコンテンツをみつけるためのベースとなるサンプルとの類似性を評価し、類似していると評価された場合に不適切なコンテンツであるとして抽出する。

10

20

30

40

50

## 【特許文献 1】

特開 2 0 0 6 - 2 9 3 4 5 5 号公報

## 【特許文献 2】

特開平 8 - 1 8 0 1 7 6 号公報

## 【特許文献 3】

特開 2 0 0 0 - 2 5 9 8 3 2 号公報

## 【特許文献 4】

特開 2 0 0 0 - 3 3 9 4 7 4 号公報

## 【発明の開示】

## 【 0 0 0 5 】

10

しかしながら、上記特許文献 1 に記載されている方法は、不適切なコンテンツをみつけるためのベースとなるコンテンツのサンプルとの類似度に基づくものであり、不適切コンテンツを抽出するために、不適切コンテンツをみつけるためのベースとなるコンテンツをサンプルとして用意する必要がある点で改善の余地を有していた。

## 【 0 0 0 6 】

特に、掲示板などのサービスで不特定多数の利用者が投稿する、画像、音声、テキスト等のデータ（コンテンツ）に対して、他人の著作権物をコピーまたは模造したいわゆる不正コンテンツまたは不法コンテンツを検出するために、予め不正画像データをサンプルとして用意し、日々掲示板などインターネットにアップされるすべての不正画像データに対応することは困難であった。

20

## 【 0 0 0 7 】

日々、多種多様なコンテンツが生成、流通する中で、著作権的に保護すべきコンテンツをすべて、予め不適切コンテンツのベースとなるコンテンツのサンプルとして用意しておくのはきわめて困難である。掲示板などのサービス母体が、コンテンツを作成していることは少ないため、事前に保護すべきコンテンツを入手することは困難であることに加え、掲示板などに投稿されるコンテンツの範囲は非常に多岐にわたっており、すべての投稿に対してもれなく不適切コンテンツをみつけるためのベースとなるサンプルとして用意することは非常に困難であるといえる。

## 【 0 0 0 8 】

本発明の目的は、上述した課題である、不正コンテンツサンプルや辞書データを予め用意する困難さを解決する不適切コンテンツ検出方法、不適切コンテンツ検出装置、そのコンピュータプログラム、およびコンテンツ公開システムを提供することにある。

30

## 【 0 0 0 9 】

本発明の第 1 の不適切コンテンツ検出方法は、不適切コンテンツ検出装置が、  
個々の利用者から投稿されたコンテンツを受け付け、  
受け付けた複数の前記投稿されたコンテンツを用いて、複数の前記投稿されたコンテンツの相互の類似度を算出し、  
前記相互の類似度に基づき、前記投稿されたコンテンツが著作権的に適切であるか否かの判定を行う。

## 【 0 0 1 0 】

40

本発明の第 2 の不適切コンテンツ検出方法は、不適切コンテンツ検出装置が、  
個々の利用者から投稿されたコンテンツを受け付け、  
受け付けた複数の前記投稿されたコンテンツを用いて、複数の前記投稿されたコンテンツ間で算出された相互の類似度に基づき、複数の前記投稿されたコンテンツが相互に類似しているか否かを判定し、類似していた場合に、相互に類似しているコンテンツ群を著作権的に不適切なコンテンツとして検出する。

## 【 0 0 1 1 】

本発明の不適切コンテンツ検出装置は、個々の利用者から投稿されたコンテンツの入力を受け付けるコンテンツ受付手段と、

受け付けた複数の前記投稿されたコンテンツを用いて、複数の前記投稿されたコンテン

50

ツの相互の類似度を算出し、前記類似度に基づき、相互に類似するコンテンツ群を検出する類似コンテンツ検出手段と、

検出された前記類似するコンテンツ群に基づき、著作権的に不適切なコンテンツを判定する不正判定手段と、を備える。

【 0 0 1 2 】

本発明のコンピュータプログラムは、コンピュータに、個々の利用者から投稿されたコンテンツから著作権的に不適切なコンテンツを検出する不適切コンテンツ検出装置を実現させるためのコンピュータプログラムであって、

前記コンピュータに、

個々の利用者から投稿されたコンテンツの入力を受け付ける手順と、

受け付けた複数の前記投稿されたコンテンツを用いて、複数の前記投稿されたコンテンツの相互の類似度を算出し、前記類似度に基づき、相互に類似するコンテンツ群を検出する手順、

検出された前記類似するコンテンツ群に基づき、著作権的に不適切なコンテンツを判定する手順、を実行させる。

【 0 0 1 3 】

本発明の第1のコンテンツ公開システムは、投稿されたコンテンツを利用者が閲覧できるように公開するコンテンツ公開システムであって、

上記不適切コンテンツ検出装置によって検出された前記不適切なコンテンツをシステム管理者に提示する提示手段と、

前記システム管理者が確認した後、前記システム管理者から削除指示を受け付ける受付手段と、

前記削除指示に従い、前記不適切なコンテンツの削除を行う削除手段と、を備える。

【 0 0 1 4 】

本発明の第2のコンテンツ公開システムは、投稿されたコンテンツを利用者が閲覧できるように公開するコンテンツ公開システムであって、

上記不適切コンテンツ検出装置によって相互に類似していると判定されたコンテンツ数が、所定数より多いか否かを判定する判定手段と、

相互に類似していると判定されたコンテンツ数が所定数より多い場合に、当該コンテンツへの利用者のアクセスを自動的に停止する制御手段と、を備える。

【 0 0 1 5 】

なお、以上の構成要素の任意の組合せ、本発明の表現を方法、装置、システム、記録媒体、コンピュータプログラムなどの間で変換したものもまた、本発明の態様として有効である。

【 0 0 1 6 】

また、本発明の各種の構成要素は、必ずしも個々に独立した存在である必要はなく、複数の構成要素が一個の部材として形成されていること、一つの構成要素が複数の部材で形成されていること、ある構成要素が他の構成要素の一部であること、ある構成要素の一部と他の構成要素の一部とが重複していること、等でもよい。

【 0 0 1 7 】

また、本発明の不適切コンテンツ検出方法およびコンピュータプログラムには複数の手順を順番に記載してあるが、その記載の順番は複数の手順を実行する順番を限定するものではない。このため、本発明の不適切コンテンツ検出方法およびコンピュータプログラムを実施するときには、その複数の手順の順番は内容的に支障しない範囲で変更することができる。

【 0 0 1 8 】

さらに、本発明の不適切コンテンツ検出方法およびコンピュータプログラムの複数の手順は個々に相違するタイミングで実行されることに限定されない。このため、ある手順の実行中に他の手順が発生すること、ある手順の実行タイミングと他の手順の実行タイミングとの一部ないし全部が重複していること、等でもよい。

10

20

30

40

50

## 【 0 0 1 9 】

本発明によれば、不正コンテンツサンプルや辞書データを予め用意することなく、投稿されるコンテンツの中から不適切なコンテンツを効率よく検出できる不適切コンテンツ検出方法、不適切コンテンツ検出装置、そのコンピュータプログラム、およびコンテンツ公開システムが提供される。

## 【 図面の簡単な説明 】

## 【 0 0 2 0 】

上述した目的、およびその他の目的、特徴および利点は、以下に述べる好適な実施の形態、およびそれに付随する以下の図面によってさらに明らかになる。

## 【 0 0 2 1 】

【 図 1 】本発明の実施の形態に係る不適切コンテンツ検出装置の構成を示すブロック図である。

【 図 2 】本発明の実施の形態における不適切コンテンツ検出の手順の一例を示すフローチャートである。

【 図 3 】本発明の実施の形態に係る不適切コンテンツ検出装置の構成を示すブロック図である。

【 図 4 】本発明の実施の形態における不適切コンテンツ検出の手順の一例を示すフローチャートである。

【 図 5 】本発明の実施の形態に係る不適切コンテンツ検出装置の構成を示すブロック図である。

【 図 6 】本発明の実施の形態における不適切コンテンツ検出の手順の一例を示すフローチャートである。

【 図 7 】本発明の実施の形態に係る確認サーバの構成を示すブロック図である。

【 図 8 】本発明の実施の形態に係るコンテンツ公開システムの構成を示す図である。

【 図 9 】本発明の不適切コンテンツ検出方法の実施例を模式的に示した図である。

【 図 1 0 】図 9 で投稿されたコンテンツの類似の状況を模式的に示した図である。

【 図 1 1 】動画の類似例を模式的に示した図である。

【 図 1 2 】静止画の類似例を模式的に示した図である。

## 【 発明を実施するための最良の形態 】

## 【 0 0 2 2 】

以下、本発明の実施の形態について、図面を用いて説明する。なお、すべての図面において、同様な構成要素には同様の符号を付し、適宜説明を省略する。

## 【 0 0 2 3 】

( 第一の実施の形態 )

図 1 は、本発明の実施の形態に係る不適切コンテンツ検出装置 1 0 0 の構成を示す機能ブロック図である。本実施形態の不適切コンテンツ検出装置 1 0 0 は、複数の投稿されたコンテンツの相互の類似度を算出し、相互の類似度に基づき、投稿されたコンテンツが著作権的に適切であるか否かの判定を行うものである。また、本実施形態において、不適切コンテンツ検出装置 1 0 0 は、投稿された複数のコンテンツの相互の類似度に基づき、複数の投稿されたコンテンツが相互に類似しているか否かを判定し、類似していた場合に、相互に類似しているコンテンツ群を著作権的に不適切なコンテンツとして検出する。

## 【 0 0 2 4 】

ここで、類似度とは、少なくとも二つのコンテンツ間が相互に類似しているか否かを示す尺度であり、たとえば、特許文献 1 の段落 0 0 1 0 および 0 0 1 1、または特許文献 4 に記載されているような公知の技術により算出できる。

## 【 0 0 2 5 】

本実施形態の不適切コンテンツ検出装置 1 0 0 は、投稿されたコンテンツの入力を受け付けるコンテンツ入力受付部 1 1 0 と、複数の投稿されたコンテンツの相互の類似度を算出し、類似度に基づき、相互に類似するコンテンツ群を検出する類似コンテンツ検出部 1 3 4 と、検出された類似するコンテンツ群に基づき、著作権的に不適切なコンテンツを判

10

20

30

40

50

定する不正判定部 1 3 6 と、を備える。

【 0 0 2 6 】

また、本実施形態の不適切コンテンツ検出装置 1 0 0 は、投稿されたコンテンツからその特徴量を抽出する特徴量抽出部 1 3 2 をさらに備え、類似コンテンツ検出部 1 3 4 は、複数の投稿されたコンテンツの特徴量を相互に照合して、特徴量の相互の類似度を算出し、相互に類似するコンテンツ群を検出する。

【 0 0 2 7 】

より詳細には、不適切コンテンツ検出装置 1 0 0 は、コンテンツ入力受付部 1 1 0 と、コンテンツ記憶部 1 2 0 と、特徴量抽出部 1 3 2 と、類似コンテンツ検出部 1 3 4 と、不正判定部 1 3 6 と、不適切コンテンツ出力部 1 4 0 と、を備えている。

10

【 0 0 2 8 】

なお、以下の各図において、本発明の本質に関わらない部分の構成については省略してある。

また、不適切コンテンツ検出装置 1 0 0 の各構成要素は、任意のコンピュータの CPU、メモリ、メモリにロードされた本図の構成要素を実現するプログラム、そのプログラムを格納するハードディスクなどの記憶ユニット、ネットワーク接続用インタフェースを中心にハードウェアとソフトウェアの任意の組合せによって実現される。そして、その実現方法、装置にはいろいろな変形例があることは、当業者には理解されるところである。以下説明する各図は、ハードウェア単位の構成ではなく、機能単位のブロックを示している。

20

【 0 0 2 9 】

コンテンツ入力受付部 1 1 0 は、個々の利用者から次々に投稿されるコンテンツの入力を受け付ける。一例として、コンテンツ入力受付部 1 1 0 は、インターネット経由で利用者より投稿されたメッセージをキャプチャして、コンテンツ部分を切り出すプログラムを搭載した CPU (Central Processing Unit) である。入力されるコンテンツとは、たとえば、掲示板などに投稿された、画像 (動画像、静止画)、音声 (音響、音楽)、テキストデータなどであり、著作権的に保護すべきコンテンツである。

【 0 0 3 0 】

入力されるコンテンツは、利用者からの投稿を直接キャプチャする形で実現されてもよいし、一度外部のサーバ (不図示) 等に蓄えられた後に逐次または一括して入力するような形であってもよい。投稿の形態は、WWWブラウザを使ってアップロードする形態であってもよいし、e-mail等にコンテンツを添付して送付するような形態であってもよい。また特定のFTPサイトに登録するような形態であってもかまわない。

30

【 0 0 3 1 】

コンテンツ記憶部 1 2 0 は、コンテンツ入力受付部 1 1 0 が受け付けたコンテンツを蓄積する。一例として、コンテンツ記憶部 1 2 0 は、ハードディスク、フラッシュメモリなどの記憶装置であり、専用の蓄積装置であっても、他の蓄積装置との兼用であってもよい。

【 0 0 3 2 】

特徴量抽出部 1 3 2 は、コンテンツ間の一致の検出、または類似度を効果的に算出するための特徴量を、コンテンツ入力受付部 1 1 0 が受け付けた各コンテンツから抽出する。一例として、特徴量抽出部 1 3 2 は、予め定められたルールにて動作するプログラムを搭載した CPU である。たとえば、コンテンツが映像情報の場合、映像中の各フレームで観測される色情報のヒストグラムを抽出するように動作してもよいし、音楽情報の場合、各時刻における周波数成分を抽出するように動作してもよい。

40

【 0 0 3 3 】

一例として、特徴量抽出部 1 3 2 は、上記特許文献 2 に記載された画像インデックス生成や、上記特許文献 3 に記載された、代表レイアウト特徴量抽出のような手順で特徴量の抽出を行う。特徴量抽出に関しては、画像や映像または音響情報などのコンテンツの相互の類似度または一致度が算出できれば、特許文献 2 や特許文献 3 に記載された特徴量でな

50

くてもかまわない。

また、特徴量抽出部 1 3 2 で抽出された特徴量は、特徴量蓄積部（不図示）に蓄積してもよいし、類似度算出毎にコンテンツ記憶部 1 2 0 に記憶されているコンテンツから抽出してもよい。

【 0 0 3 4 】

類似コンテンツ検出部 1 3 4 は、コンテンツ間の類似度判定または一致検出を行うために、特徴量抽出部 1 3 2 により抽出された特徴量の相互の類似度を算出し、予め定められた特定閾値以上の類似度を持つコンテンツ群を検出する。検出した結果は類似コンテンツ情報として一時記憶部（不図示）に記憶し、後述する不正判定部 1 3 6 に出力する。一例として、類似コンテンツ検出部 1 3 4 は、予め定められたルールにて動作するプログラムを搭載した CPU である。専用の CPU であってもよいし、他の CPU との兼用であってもよい。

10

【 0 0 3 5 】

類似するコンテンツの検出方法として、たとえば、特許文献 3 に記載された、代表レイアウト特徴量に基づく類似コンテンツ検出を用いる。コンテンツの一致に際しては、映像中の一部または全部のフレームの一致または類似や、静止画像の一部または全体の一致または類似、音響の一部または全体の一致または類似を検出できれば、特許文献 3 以外の方法で検出を行ってもよい。

【 0 0 3 6 】

すなわち、本実施形態の不適切コンテンツ検出装置 1 0 0 において、類似コンテンツ検出部 1 3 4 は、投稿されたコンテンツが映像であった際に、映像に含まれる一部または全部のフレーム群が相互に類似しているときに、コンテンツ間の類似度が閾値より高いとすることができる。また、類似コンテンツ検出部 1 3 4 は、投稿されたコンテンツが静止画像であった際に、静止画像の一部または全体が類似しているときに、コンテンツ間の類似度が閾値より高いとすることができる。さらに、類似コンテンツ検出部 1 3 4 は、投稿されたコンテンツが音響または音楽を含む際に、音響または音楽の一部または全体のフレーズが類似しているときに、コンテンツ間の類似度が閾値より高いとすることができる。

20

【 0 0 3 7 】

不正判定部 1 3 6 は、類似コンテンツ検出部 1 3 4 にて検出された類似コンテンツ情報に基づいて、コンテンツの不正を判定する。一例として、不正判定部 1 3 6 は、予め定められたルールにて動作するプログラムを搭載した CPU である。専用の CPU であってもよいし、他の CPU との兼用であってもよい。

30

【 0 0 3 8 】

不適切コンテンツ出力部 1 4 0 は、不正判定部 1 3 6 における不正判定結果をうけて、不適切なコンテンツを出力する。一例として、不適切コンテンツ出力部 1 4 0 は、予め定められたルールにて動作するプログラムを搭載した CPU である。専用の CPU であってもよいし、他の CPU との兼用であってもよい。例として、不適切コンテンツ出力部 1 4 0 は、不正判定部 1 3 6 で不適切なコンテンツと判定された不適切なコンテンツをリストアップした不適切コンテンツリストを作成し、出力するリスト作成部（不図示）を含むこともできる。

40

【 0 0 3 9 】

上記不適切コンテンツ検出装置 1 0 0 の各種ユニット（コンテンツ入力受付部 1 1 0 ~ 不適切コンテンツ出力部 1 4 0 ）を各種機能として実現させるためのコンピュータプログラムは、上述の CPU が利用するメモリ（不図示）に記憶され、CPU により実行される。

【 0 0 4 0 】

本実施形態のコンピュータプログラムは、コンピュータに、投稿されたコンテンツから著作権的に不適切なコンテンツを検出する不適切コンテンツ検出装置 1 0 0 を実現させるためのコンピュータプログラムであって、コンピュータに、投稿されたコンテンツの入力を受け付ける手順と、投稿されたコンテンツからその特徴量を抽出する手順、複数の投稿

50

されたコンテンツの特徴量を相互に照合して、特徴量の相互の類似度を算出する手順、算出された類似度に基づいて、相互に類似するコンテンツ群を検出する手順、検出された類似するコンテンツ群に基づき、著作権的に不適切なコンテンツを判定する手順、を実行させるように記述されている。

【0041】

コンピュータプログラムは、たとえば、コンピュータ（CPU）が読取可能な記録媒体（メモリ）に格納される。例として、記録媒体は、PROM（Programmable Read Only Memory）、ハードディスク、DVD-ROM、CD-ROM、FDなどである。

【0042】

このように構成された本実施形態の不適切コンテンツ検出装置100の動作について、以下に説明する。図2は、本実施形態の不適切コンテンツ検出装置100の動作の一例を示すフローチャートである。以下、図1および図2を用いて説明する。

10

【0043】

はじめに、コンテンツ入力受付部110により利用者により投稿された複数のコンテンツの入力が受け付けられる（S11）。コンテンツは、逐次入力されてもよいし、複数まとめて入力されてもよい。受け付けられたコンテンツがコンテンツ記憶部120に記憶され（S13）、特徴量抽出部132が、コンテンツ入力受付部110が受け付けた各コンテンツから類似度算出のための特徴量を抽出する（S15）。

【0044】

そして、類似コンテンツ検出部134が、受け付けたコンテンツの中から順次コンテンツ対を選択し、（S17）、選択されたコンテンツ対の類似度を算出し（S19）、類似度が予め定めた閾値以上であるかを判定する（S21）。類似度が予め定めた閾値以上であった場合に（S21のYes）、類似コンテンツとして検出し、検出されたコンテンツ対を相互類似コンテンツ情報として一時記憶部に記憶する（S23）。

20

【0045】

ステップS17からステップS23までを、入力コンテンツのすべての組合せに対して行い（S25）、それぞれ相互に類似するコンテンツを検出する。

【0046】

そして、すべての組み合わせについて類似コンテンツの検出が終了したとき（S25のYes）、不正判定部136により、ステップS23で記憶した相互類似コンテンツ情報に基づいて不正判定が行われる（S27）。一例として、特定の相互に類似するコンテンツが予め定められた件数以上検出された場合、それらの類似するコンテンツ群に対して不正と判定する。一例として、予め定められた件数が2件の場合、ステップS23にて記憶されたコンテンツはすべて、不適切コンテンツと判定される。予め定められた件数がN件の場合、ステップS21で検出されたコンテンツ対の接続関係に基づき、N件以上の相互接続を検出して検出されたコンテンツを不適切コンテンツと判定する。

30

【0047】

そして、不適切コンテンツ出力部140により、不正と判定されたコンテンツの出力が行われる（S29）。

【0048】

以上説明したように、本実施形態の不適切コンテンツ検出装置100によれば、不適切なコンテンツをみつけるためのベースとなるサンプルコンテンツ（辞書データ）を用意することなく、著作権的に不適切なコンテンツを効率よく自動的に検出するという、発明の目的を達成することができる。

40

また、著作権的な適切度の判定を効率よく行うことができる。そして、掲示板などに不適切なコンテンツが公開されるのを防ぐことが可能となる。

【0049】

権利侵害をしている不適切コンテンツでは、テレビ録画、雑誌、CD、DVDなどのオリジナルコンテンツが存在し、オリジナルコンテンツを一部またはそのまま投稿したり、他者が投稿したコンテンツをダウンロードしてそのまま転用したりするという特性をもつ

50

ため重複が多い。投稿されたコンテンツ間で類似度評価を行い、投稿された動画、静止画、音響の一致または重複を検知し、重複の多いコンテンツを著作権的に不適切なコンテンツとして検出することにより、不適切コンテンツのサンプルを予め用意することなく、著作権的な適切度の判定および不適切なコンテンツの検出ができるようになり、本発明の目的を達成することができる。

#### 【0050】

(第二の実施の形態)

図3は、本発明の実施の形態に係る不適切コンテンツ検出装置200の構成を示す機能ブロック図である。本実施形態の不適切コンテンツ検出装置200は、上記実施形態の不適切コンテンツ検出装置100とは、コンテンツを所定の条件で選択し、選択されたコンテンツについて類似度評価を行う点で相違する。

10

#### 【0051】

本実施形態の不適切コンテンツ検出装置200において、類似コンテンツ検出部234は、所定の期間内に投稿されたコンテンツのうち、所定数以上の投稿されたコンテンツ間で相互に類似度が閾値より高いコンテンツを相互に類似するコンテンツ群として検出し、不正判定部236は、検出された類似するコンテンツ群を不適切なコンテンツと判定する。

#### 【0052】

また、本実施形態の不適切コンテンツ検出装置200において、類似コンテンツ検出部234は、複数の投稿されたコンテンツから、所定の基準でコンテンツを選択した後に、選択されたコンテンツ間で相互に類似度を算出し、所定数以上のコンテンツ間で、相互に類似度が閾値より高いコンテンツ群を検出し、不正判定部236は、検出された前記コンテンツ群を、不適切なコンテンツと判定する。本実施形態の不適切コンテンツ検出装置200は、たとえば、所定期間に投稿されたコンテンツについて類似度評価を行う。

20

#### 【0053】

ここで、所定の基準とは、コンテンツが投稿された日時、コンテンツを投稿した投稿者の国籍、投稿されたコンテンツのサイズなどを含む。これらの情報は、コンテンツ投稿記録などに記録されており、コンテンツ投稿記録、すなわち、投稿時刻、投稿者の国籍、コンテンツのサイズなどに基づいてコンテンツを選択し、選択されたコンテンツ間で類似度を算出し、不適切コンテンツを検出することができる。

30

#### 【0054】

本実施形態の不適切コンテンツ検出装置200は、特徴量抽出部232によって抽出された特徴量を記憶する特徴量蓄積部252と、特徴量蓄積部252から、所定の基準で選択されたコンテンツの特徴量を取得する蓄積特徴量選択部254と、を備え、類似コンテンツ検出部234は、蓄積特徴量選択部254によって選択されたコンテンツの特徴量の相互の類似度を算出し、相互に類似するコンテンツを検出する。

#### 【0055】

さらに、本実施形態の不適切コンテンツ検出装置200において、特徴量蓄積部252は、投稿毎に投稿されたコンテンツの特徴量を蓄積しておき、類似コンテンツ検出部234は、新たに投稿された新規コンテンツと複数の蓄積されたコンテンツの特徴量に基づいて、複数の蓄積されたコンテンツ毎に、新規コンテンツとの類似度を算出し、新規コンテンツとの類似度が閾値より高いコンテンツが存在した場合に、新規コンテンツおよび、新規コンテンツとの類似度が閾値より高いコンテンツのコンテンツ群を検出し、不正判定部236は、検出されたコンテンツ群を不適切なコンテンツと判定する。

40

#### 【0056】

詳細には、本実施形態の不適切コンテンツ検出装置200は、上記実施形態の不適切コンテンツ検出装置100と同じコンテンツ入力受付部110と、コンテンツ記憶部120と、不適切コンテンツ出力部140と、に加え、特徴量抽出部232と、類似コンテンツ検出部234と、不正判定部236と、特徴量蓄積部252と、蓄積特徴量選択部254と、を備えている。

50

## 【 0 0 5 7 】

特徴量抽出部 2 3 2 は、コンテンツ間の一致または類似度を効果的に算出するための特徴量をコンテンツから抽出する。抽出された特徴量は特徴量蓄積部 2 5 2 に蓄積される。特徴量抽出部 2 3 2 は、たとえば、コンテンツが映像情報の場合、映像中の各フレームで観測される色情報のヒストグラムを抽出するように動作してもよいし、音楽情報の場合、各時刻における周波数成分を抽出するように動作してもよい。

## 【 0 0 5 8 】

一例として、特徴量抽出部 2 3 2 は、上記実施形態で記載した特許文献 2 に記載された画像インデックス生成や、特許文献 3 に記載された、代表レイアウト特徴量抽出のような手順で特徴量の抽出を行う。特徴量抽出に関しては、画像や映像または音響情報などのコンテンツの相互の類似度または一致度が算出できれば、特許文献 2 や特許文献 3 に記載された特徴量でなくともかまわない。

10

## 【 0 0 5 9 】

特徴量蓄積部 2 5 2 は、特徴量抽出部 2 3 2 にて抽出された特徴量を蓄積する。一例として、ハードディスクやフラッシュメモリなどの記憶装置があり、専用の蓄積装置であっても、他の蓄積装置との兼用であってもよい。

## 【 0 0 6 0 】

蓄積特徴量選択部 2 5 4 は、特徴量蓄積部 2 5 2 から類似度を算出すべき特徴量を選択する。一例として、予め定められたルールにて動作するプログラムを搭載した CPU である。専用の CPU であってもよいし、他の CPU との兼用であってもよい。一例として、蓄積特徴量選択部 2 5 4 は、特徴量蓄積部 2 5 2 に蓄積された特徴量のうち、予め定められた期間に投稿されたコンテンツの特徴量を、投稿記録の投稿日時に基づき選択する。このほか、投稿記録に記録されているコンテンツの投稿者や投稿者の国籍、投稿されたコンテンツのサイズなどに応じて、類似度評価するコンテンツを選択してもかまわない。

20

## 【 0 0 6 1 】

類似コンテンツ検出部 2 3 4 は、受け付けたコンテンツの特徴量または蓄積されたコンテンツの特徴量を相互に照合することで、相互の類似度を算出し、類似度に基づいて類似コンテンツを検出する。一例として、予め定められたルールにて動作するプログラムを搭載した CPU である。専用の CPU であってもよいし、他の CPU との兼用であってもよい。類似コンテンツグループ（コンテンツ群）検出に際しては、上記実施形態で述べたようにすべての特徴量対の類似度を総当り的にもとめる方法（図 2 のステップ S 1 7 ~ S 2 5）を用いてもよいし、まず、第一に、新規に投稿された入力コンテンツの特徴量と、蓄積された特徴量との類似度の算出を行い、類似コンテンツが検出されたときのみ、相互に類似度算出を行うという方法を用いてもよい。

30

## 【 0 0 6 2 】

すなわち、本実施形態の不適切コンテンツ検出装置 2 0 0 において、特徴量蓄積部 2 5 2 は、これまでに投稿されたコンテンツの特徴量を蓄積しておき、類似コンテンツ検出部 2 3 4 は、新たに投稿された新規コンテンツと複数の蓄積されたコンテンツの特徴量に基づいて、複数の蓄積されたコンテンツ毎に、新規コンテンツとの類似度を算出し、新規コンテンツとの類似度が閾値より高いコンテンツが存在した場合に、蓄積されたコンテンツに新規コンテンツを加えて、相互の類似度を算出し、所定数以上のコンテンツ間で相互に類似度が閾値より高いコンテンツ群を検出してもよい。

40

すなわち、新規コンテンツとの類似度が閾値より高いコンテンツが存在しない場合は、コンテンツ群の検出を行わなくてよい。

## 【 0 0 6 3 】

また、このとき、類似コンテンツ検出部 2 3 4 は、新規コンテンツと比較する蓄積されたコンテンツとして、所定期間に投稿されたコンテンツのみを用いてもよい。

## 【 0 0 6 4 】

不正判定部 2 3 6 は、類似コンテンツ検出部 2 3 4 にて検出された類似コンテンツ情報をもとにコンテンツの不正を判定する。本実施形態において、不正判定部 2 3 6 は、類似

50

コンテンツ検出部 234 にて検出された類似コンテンツグループが予め定めた数以上であった場合、検出された類似コンテンツグループを不適切なコンテンツと判定する。一例として、不正判定部 236 は、予め定められたルールにて動作するプログラムを搭載した CPU である。専用の CPU であってもよいし、他の CPU との兼用であってもよい。

【0065】

上記不適切コンテンツ検出装置 200 の各種ユニット（コンテンツ入力受付部 110 ~ 不適切コンテンツ出力部 140、特徴量抽出部 232 ~ 蓄積特徴量選択部 254）を各種機能として実現させるためのコンピュータプログラムは、上述の CPU が利用するメモリ（不図示）に記憶され、CPU により実行される。

【0066】

本実施形態のコンピュータプログラムは、上記実施形態のコンピュータプログラムの手順に加え、コンピュータに、所定の期間内に投稿されたコンテンツのうち、所定数以上の投稿されたコンテンツ間で相互に類似度が閾値より高いコンテンツを相互に類似するコンテンツ群として検出する手順をさらに実行させるように記述されている。

【0067】

このように構成された本実施形態の不適切コンテンツ検出装置 200 の動作について以下に説明する。図 4 は、本実施形態の不適切コンテンツ検出装置 200 の動作の一例を示すフローチャートである。以下、図 3 および図 4 を用いて説明する。

【0068】

なお、本実施形態の不適切コンテンツ検出装置 200 の動作において、図 2 のフローチャートのステップ S11、S13、および S29 は同じであるので、説明は省略する。

【0069】

特徴量抽出部 232 により、ステップ S11 で入力された各コンテンツから類似度算出のための特徴量が抽出される（S31）。抽出された特徴量は、特徴量蓄積部 252 に蓄積される。そして、蓄積特徴量選択部 254 により、予め特徴量蓄積部 252 に記憶された特徴量から相互類似度評価用に利用すべき参照コンテンツの特徴量が選択される（S33）。たとえば、上述したように投稿記録に基づき、所定期間に投稿されたコンテンツの特徴量を選択する。

【0070】

そして、類似コンテンツ検出部 234 により、ステップ S33 で選択された特徴量に、新規に入力されたコンテンツの特徴量を加えた特徴量間で相互に類似度を算出して、類似コンテンツグループの検出を行う（S35）。

【0071】

ステップ S35 にて検出された類似コンテンツグループが予め定めた数以上であった場合、不正判定部 236 により、検出された類似コンテンツグループが不正コンテンツと判定され（S37）、不適切コンテンツ出力部 140 により不正と判定したコンテンツが出力される（S29）。

【0072】

以上説明したように、本実施形態の不適切コンテンツ検出装置 200 によれば、特定期間内に数多く重複して投稿された、オリジナルコンテンツの存在する、コピーとおぼしき投稿を検出することができ、不適切なコンテンツをみつけるためのベースとなるサンプルコンテンツ（辞書データ）を用意することなく、著作権的に不適切なコンテンツを自動的に検出するという、発明の目的を達成することができる。

【0073】

（第三の実施の形態）

図 5 は、本発明の実施の形態に係る不適切コンテンツ検出装置 300 の構成を示す機能ブロック図である。本実施形態の不適切コンテンツ検出装置 300 は、上記実施形態の不適切コンテンツ検出装置 100 および不適切コンテンツ検出装置 200 とは、検出された不適切なコンテンツを辞書として登録し、辞書との照合を行うことで不適切なコンテンツの検出を行う点、および不適切なコンテンツが検出された場合、管理者などにアラーム提

10

20

30

40

50

示する点で相違する。

【 0 0 7 4 】

本実施形態の不適切コンテンツ検出装置 3 0 0 は、不正判定部 2 3 6 で不適切と判定された不適切なコンテンツまたは該不適切なコンテンツの特徴量を不適切コンテンツ辞書データとして不適切コンテンツ特徴量蓄積部 3 1 2 に記憶する不適切コンテンツ特徴量登録部 3 1 4 と、新規に投稿されたコンテンツに対して、不適切コンテンツ辞書データと照合を行うことにより不適切なコンテンツの検出を行う類似コンテンツ検出部 3 3 4 と、をさらに備える。

【 0 0 7 5 】

詳細には、本実施形態の不適切コンテンツ検出装置 3 0 0 は、上記実施形態の不適切コンテンツ検出装置 1 0 0 と同じコンテンツ入力受付部 1 1 0 と、コンテンツ記憶部 1 2 0 と、不適切コンテンツ出力部 1 4 0 と、ならびに、上記実施形態の不適切コンテンツ検出装置 2 0 0 と同じ特徴量抽出部 2 3 2 と、不正判定部 2 3 6 と、特徴量蓄積部 2 5 2 と、に加え、類似コンテンツ検出部 3 3 4 と、不適切コンテンツ特徴量蓄積部 3 1 2 と、不適切コンテンツ特徴量登録部 3 1 4 と、アラーム提示部 3 2 0 と、を備えている。

10

【 0 0 7 6 】

不適切コンテンツ特徴量蓄積部 3 1 2 は、既に不適切なコンテンツであると判定されたコンテンツの特徴量を蓄積する。一例として、ハードディスクやフラッシュメモリなどの記憶装置があり、専用の蓄積装置であっても、他の蓄積装置との兼用であってもよい。

【 0 0 7 7 】

不適切コンテンツ特徴量登録部 3 1 4 は、不適切なコンテンツであると判定されたコンテンツの特徴量を不適切コンテンツ特徴量蓄積部 3 1 2 に登録する。一例として予め定められたルールにて動作するプログラムを搭載した CPU である。たとえば、不正コンテンツと判定されたコンテンツの特徴量を特徴量蓄積部 2 5 2 から、不適切コンテンツ特徴量蓄積部 3 1 2 に変更するという動作を行う。また、放送局などから提供された著作物コンテンツの特徴量を新たに登録するように動作してもよい。

20

【 0 0 7 8 】

類似コンテンツ検出部 3 3 4 は、入力されたコンテンツの特徴量または蓄積されたコンテンツの特徴量の相互の類似度を算出し、類似コンテンツを検出する。一例として、予め定められたルールにて動作するプログラムを搭載した CPU である。専用の CPU であってもよいし、他の CPU との兼用であってもよい。特徴量間の類似度の算出方法として、たとえば上記実施形態にて記述した特許文献 3 に記載された、代表レイアウト特徴量に基づく類似コンテンツ検出を用いる。コンテンツの一致に際しては、映像中の一部または全部のフレームの一致または類似や、静止画像の一部または全体の一致または類似、音響の一部または全体の一致または類似を検出できれば、特許文献 3 以外の方法で検出を行ってもよい。

30

【 0 0 7 9 】

アラーム提示部 3 2 0 は、不適切なコンテンツが検出された際にコンテンツ管理者などに対してアラームを提示する。一例として、アラーム用のテキストまたは映像を提示するためのモニタ、音響を出力するためのスピーカなどが上げられる。

40

【 0 0 8 0 】

上記不適切コンテンツ検出装置 3 0 0 の各種ユニット（コンテンツ入力受付部 1 1 0 ～不適切コンテンツ出力部 1 4 0、特徴量抽出部 2 3 2 ～特徴量蓄積部 2 5 2、不適切コンテンツ特徴量蓄積部 3 1 2 ～類似コンテンツ検出部 3 3 4）を各種機能として実現させるためのコンピュータプログラムは、上述の CPU が利用するメモリ（不図示）に記憶され、CPU により実行される。

【 0 0 8 1 】

本実施形態のコンピュータプログラムは、上記実施形態のコンピュータプログラムの手順に加え、コンピュータに、不適切なコンテンツを判定する手順で不適切と判定された不適切なコンテンツまたは該不適切なコンテンツの特徴量を不適切コンテンツ辞書データと

50

して不適切コンテンツ特徴量蓄積部 3 1 2 に記憶する手順、新規に投稿されたコンテンツに対して、不適切コンテンツ辞書データと照合を行うことにより不適切なコンテンツの検出を行う手順、をさらに実行させるように記述されている。

【 0 0 8 2 】

このように構成された本実施形態の不適切コンテンツ検出装置 3 0 0 の動作について、以下に説明する。図 6 は、本実施形態の不適切コンテンツ検出装置 3 0 0 の動作の一例を示すフローチャートである。以下、図 5 および図 6 を用いて説明する。

【 0 0 8 3 】

なお、本実施形態の不適切コンテンツ検出装置 3 0 0 の動作において、図 2 のフローチャートのステップ S 1 1 および S 1 3 は同じであるので、説明は省略する。

【 0 0 8 4 】

特徴量抽出部 2 3 2 により、ステップ S 1 1 で入力された各コンテンツから類似度算出のための特徴量が抽出される ( S 4 1 )。抽出された特徴量は、特徴量蓄積部 2 5 2 に蓄積される。そして、類似コンテンツ検出部 3 3 4 により、入力コンテンツの特徴量と、不適切コンテンツ特徴量蓄積部 3 1 2 に蓄積された各コンテンツの特徴量との間の類似度が算出される ( S 4 3 )。

【 0 0 8 5 】

ステップ S 4 3 にて算出された類似度が予め定めた閾値以上であった場合 ( S 4 5 の Y e s )、不正判定部 2 3 6 により、類似度が閾値以上の特徴量の入力コンテンツが不適切なコンテンツと判定され ( S 5 1 )、不適切コンテンツ出力部 1 4 0 により不正と判定したコンテンツが出力される ( S 5 3 )。そして、アラーム提示部 3 2 0 が、情報管理者に不適切コンテンツ検出のアラームをあげる ( S 5 5 )。そして、不適切コンテンツ特徴量登録部 3 1 4 により、新たに不適切コンテンツと判定されたコンテンツの特徴量が不適切コンテンツ特徴量蓄積部 3 1 2 に登録される ( S 5 7 )。

【 0 0 8 6 】

ステップ S 4 3 にて算出されたコンテンツが、不適切コンテンツ特徴量蓄積部 3 1 2 に蓄積されたすべてのコンテンツに対して予め定めた閾値未満であった場合 ( S 4 5 の N o ) は、類似コンテンツ検出部 3 3 4 により、入力されたコンテンツの特徴量ならびに、特徴量蓄積部 2 5 2 に蓄積された特徴量で特徴量対が選択され ( S 6 1 )、選択された特徴量間の類似度が算出される ( S 6 3 )。そして、類似コンテンツ検出部 3 3 4 により、類似度が予め定めた閾値以上か否かが判定される ( S 6 5 )。類似度が予め定めた閾値以上であった場合 ( S 6 5 の Y e s ) に、類似コンテンツ検出部 3 3 4 により類似特徴量対として検出され、類似情報として一時記憶部に記憶する ( S 6 7 )。これを入力コンテンツからの特徴量と、特徴量蓄積部 2 5 2 に蓄積された特徴量をあわせたすべての特徴量の組合せに対して行い ( S 6 9 )、それぞれ相互に類似するコンテンツの特徴量を抽出する。

【 0 0 8 7 】

そして、すべての組み合わせについて類似判定が終了したとき ( S 6 9 の Y e s )、不正判定部 2 3 6 により、特徴量間の類似情報から不正判定が行われる ( S 7 1 )。一例として、特定の相互に類似するコンテンツの特徴量が、入力コンテンツの特徴量が加わることにより、予め定められた件数以上検出された場合、入力コンテンツならびに、入力コンテンツと相互に類似する特徴量蓄積部 2 5 2 に蓄積された特徴量に対応するコンテンツ記憶部 1 2 0 に蓄積されたコンテンツを不正と判定する。一例として、予め定められた件数が 2 件の場合、ステップ S 6 5 にて新たに検出された入力コンテンツを含むコンテンツはすべて、不適切コンテンツと判定される。予め定められた件数が N 件の場合、ステップ S 6 5 で検出されたコンテンツの特徴量対の接続関係に基づき、N 件以上の相互接続を検出して検出された特徴量に対応するコンテンツが不適切コンテンツと判定される。

【 0 0 8 8 】

そして、不適切コンテンツ出力部 1 4 0 により、不正と判定されたコンテンツの出力が行われる ( S 7 3 )。さらに、アラーム提示部 3 2 0 が、情報管理者に不適切コンテンツ検出のアラームをあげる ( S 7 5 )。そして、不適切コンテンツ特徴量登録部 3 1 4 によ

10

20

30

40

50

り、新たに不適切コンテンツと判定されたコンテンツの特徴量が、特徴量蓄積部 2 5 2 から消去され、入力コンテンツの特徴量ならびに新たに不適切コンテンツと判定したコンテンツの特徴量が不適切コンテンツ特徴量蓄積部 3 1 2 に登録される ( S 7 7 ) 。

【 0 0 8 9 】

なお、不適切コンテンツ検出装置 3 0 0 において、所定期間または所定の利用者または所定のコンテンツに対する相互の類似度の算出により不適切コンテンツ辞書データを生成する生成部 ( 類似コンテンツ検出部 3 3 4、不正判定部 2 3 6 ) をさらに備えてもよい。これにより、所定期間または所定の利用者など、特定のグループ毎に辞書データを生成することができる。

また、この生成部は、相互の類似度の算出による不適切コンテンツ検出を定期的に行い、不適切コンテンツ辞書データの生成または更新を行ってもよい。

10

【 0 0 9 0 】

また、特徴量蓄積部 2 5 2 に蓄積する特徴量を、特定期間において投稿されたコンテンツの特徴量に制限するようにすることにより、特定期間内に投稿されたコンテンツの相互類似度を算出するようにすることもできる。また、不適切コンテンツ特徴量蓄積部 3 1 2 に、既に不適切コンテンツとして検出対象である、著作権保護コンテンツの特徴量を合わせて登録してもよい。

【 0 0 9 1 】

また、不適切コンテンツ特徴量登録部 3 1 4 は、特徴量蓄積部 2 5 2 からの消去や不適切コンテンツ特徴量蓄積部 3 1 2 への登録を自動的に行う代わりに、コンテンツ管理者等が確認をしてから登録を行うように変更してもよい。

20

【 0 0 9 2 】

すなわち、上記不適切コンテンツ検出装置 3 0 0 において、不適切コンテンツ特徴量登録部 3 1 4 によって生成された不適切コンテンツ辞書データを情報管理者に提示し、該情報管理者が不適切コンテンツ辞書データを確認し、保持する不適切コンテンツ辞書データの指示を受け付け、指示された不適切コンテンツ辞書データのみを不適切コンテンツ特徴量蓄積部 3 1 2 に記憶してもよい。

【 0 0 9 3 】

相互類似性を判定する類似度の閾値や類似コンテンツは判定する場合の件数、または使用する特徴量をジャンルや時期、投稿者などの状況に応じて適応的に変更したり、情報管理者が手動で調整したりできるような機構があってもかまわない。

30

【 0 0 9 4 】

また、不適切コンテンツ検出装置 3 0 0 は、所定の利用者または所定のコンテンツを予め登録する登録部 ( 不図示 ) を備えてもよい。このとき、不正判定部 2 3 6 は、所定の利用者により投稿されたコンテンツまたは所定のコンテンツは不適切コンテンツと判定しなくてもよい。

【 0 0 9 5 】

以上説明したように、本実施形態の不適切コンテンツ検出装置 3 0 0 によれば、所定の期間内に投稿されたコンテンツ内で、類似度の高いコンテンツが所定数以上検出されたときにそれらのコンテンツ群を不適切コンテンツとするので、特定期間内に数多く重複して投稿された、オリジナルコンテンツの存在する、コピーとおぼしき投稿を検出することができ、不適切コンテンツをみつけるためのベースとなるサンプルコンテンツ ( 辞書データ ) を用意することなく、著作権的に不適切コンテンツを自動的に効率よく検出するという、発明の目的を達成することができる。

40

【 0 0 9 6 】

権利侵害をしている不適切コンテンツでは、テレビ録画、雑誌、CD、DVDなどのオリジナルコンテンツが存在し、オリジナルコンテンツを一部またはそのまま投稿したり、他者が投稿したコンテンツをダウンロードしてそのまま転用したり、短期間に連続的に投稿されるという特性をもつため短期間での重複が多い。本実施形態によれば、予め定めた特定期間に投稿されたコンテンツ間で類似度評価を行い、投稿された動画、静止画、音響

50

の一致または重複を検知し、予め定めた数以上の重複の多いコンテンツを著作権的に不適切なコンテンツとして検出することにより、不適切コンテンツのサンプルを予め用意することなく、不適切なコンテンツを検出できるようになり、本発明の目的を達成することができる。

【0097】

なお、本実施形態では、不適切コンテンツ特徴量蓄積部312に不適切なコンテンツの特徴量のみを蓄積し、辞書データとして準備するので、辞書データとして不正コンテンツのサンプルを用意する場合に比べて、その容量は非常に小さくて済む。

【0098】

さらに、本実施形態の不適切コンテンツ検出装置300によれば、不適切なコンテンツが検出されたとき、アラームをあげることで情報管理者の確認を促すことができ、これにより不適切なコンテンツが公開されることを防ぐことができる。

10

【0099】

(第四の実施の形態)

図7は、本発明の実施の形態に係る確認サーバ400の構成を示す機能ブロック図である。

本実施形態の確認サーバ400は、通信部410と、アラーム提示部420と、管理者確認および登録部430と、蓄積部440と、制御部450と、を備えている。

【0100】

通信部410は、ネットワーク402を通じて端末(不図示)や他の管理装置(不図示)とコンテンツや伝達情報のやり取りを行う。通信部410は、一例として、ネットワーク402を介した通信を行う専用ボードなどがあげられる。

20

【0101】

アラーム提示部420は、コンテンツ管理者に不適切コンテンツ検出のアラーム情報を提示する。アラーム提示部420は、一例として、アラーム用のテキストまたは映像を提示するためのモニタ、音情報を出力するためのスピーカなどがあげられる。あるいは、アラーム情報をリストアップしたリストを作成し、印字出力するプリンタなどでもよい。

【0102】

管理者確認および登録部430は、管理者が、不適切コンテンツ検出結果を確認したり、不適切コンテンツを新たに登録するためのものである。管理者確認および登録部430は、一例として結果を表示するモニタとキーボード、タッチパネルなどの入力機器の組合せがあげられる。あるいは、不適切コンテンツ検出結果をリストアップしたリストを作成し、印字出力するプリンタなどでもよい。

30

【0103】

蓄積部440は、コンテンツや抽出した特徴量、不適切コンテンツの特徴量などを蓄積する。蓄積部440は、一例として、ハードディスクやフラッシュメモリなどの記憶装置があり、専用の蓄積装置であっても、他の蓄積装置との兼用であってもかまわない。

【0104】

制御部450は、確認サーバ400の各要素および装置全体を制御するとともに、コンテンツ特徴量の抽出や照合など、不適切コンテンツの抽出を行う。制御部450は、一例として、プログラムを搭載したCPUである。

40

【0105】

上記通信部410、アラーム提示部420、管理者確認および登録部430、蓄積部440、制御部450をあわせて、通信機能、蓄積機能、モニタ機能、入力機能を備えたコンピュータにて構築可能である。

【0106】

制御部450は、通信部410からコンテンツをうけとり入力するコンテンツ入力受付部452と、不適切コンテンツ検出結果を出力する不適切コンテンツ出力部454と、不適切コンテンツであると判定されたコンテンツの特徴量を不適切コンテンツ特徴量蓄積部444に登録する不適切コンテンツ特徴量登録部456と、蓄積部440に格納されてい

50

るコンテンツ、特徴量、不適切コンテンツ特徴量を管理する蓄積管理部 4 6 0 と、コンテンツの特徴量をコンテンツから抽出する特徴量抽出部 4 6 2 と、特徴量間の類似度算出により類似コンテンツを抽出する類似コンテンツ検出部 4 6 4 と、類似コンテンツ検出部 4 6 4 の検出結果を基に不適切なコンテンツを判定する不正判定部 4 6 6 を含む。

【 0 1 0 7 】

蓄積部 4 4 0 は、入力されたコンテンツを蓄積するコンテンツ記憶部 4 4 2 と、抽出されたコンテンツ特徴量を蓄積する特徴量蓄積部 4 4 4 と、不適切コンテンツと判定されたコンテンツの特徴量を蓄積する不適切コンテンツ特徴量蓄積部 4 4 6 と、を含む。

【 0 1 0 8 】

このように構成された本実施形態の確認サーバ 4 0 0 の動作については、上記実施形態の不適切コンテンツ検出装置 3 0 0 と同様であるので説明は省略する。

【 0 1 0 9 】

本実施形態の確認サーバ 4 0 0 によれば、上記実施形態の不適切コンテンツ検出装置 3 0 0 と同様な効果を奏する。

【 0 1 1 0 】

( 第五の実施の形態 )

図 8 は、本発明の実施の形態に係るコンテンツ公開システム 5 0 0 の構成を示す図である。

本実施形態のコンテンツ公開システム 5 0 0 は、映像公開サービスを提供するシステムであり、ネットワーク 5 0 2 と、ネットワーク 5 0 2 上に接続された複数の端末装置 5 1 0 と、不適切なコンテンツの検出および確認を行う確認サーバ 5 2 0 と、利用者により投稿された公開データを記憶する公開データ記憶部 5 3 0 と、公開データを公開するための公開サーバ 5 4 0 と、確認サーバ 5 2 0 が不適切なコンテンツを検出した際に、公開データ記憶部 5 3 0 から公開データを削除する公開データ削除部 5 5 0 と、を備える。

【 0 1 1 1 】

端末装置 5 1 0 は、ネットワーク 5 0 2 へのアクセス機能を有する端末装置であり、汎用のコンピュータで実現可能である。

ネットワーク 5 0 2 は、データの送受信が可能なインターネット等の通信網である。

【 0 1 1 2 】

確認サーバ 5 2 0 は、ネットワーク 5 0 2 を解して端末装置 5 1 0 から送信された投稿データを受け付けて、公開データ記憶部 5 3 0 に格納する機能を有するとともに、公開データ記憶部 5 3 0 に記憶された投稿データのうちで、不適切なものを検出して、削除する機能を有するサーバコンピュータである。確認サーバ 5 2 0 は、図 7 の上記実施形態で述べた確認サーバ 4 0 0 により実現可能である。不適切コンテンツ出力部 5 2 2 は、図 7 における確認サーバ 4 0 0 の不適切コンテンツ出力部 4 5 4 と同様な構成により実現可能である。

【 0 1 1 3 】

公開サーバ 5 4 0 は、公開データ記憶部 5 3 0 に記憶されているデータを、ネットワーク 5 0 2 を通じて公開する機能を有するサーバコンピュータであり、通常の WWW サーバとしての機能を有するコンピュータにより実現可能である。

【 0 1 1 4 】

公開データ削除部 5 5 0 は、確認サーバ 5 2 0 の検出結果を受けて、公開データ記憶部 5 3 0 中の公開データを削除する。一例として、予め定められたルールにて動作するプログラムを搭載した CPU である。専用の CPU であってもよいし、他の CPU との兼用であってもよい。

【 0 1 1 5 】

以上説明したように本実施形態のコンテンツ公開システム 5 0 0 によれば、不適切コンテンツが検出された際に、公開データを削除する機構が実現できるため、ネットワーク 5 0 2 を介してアクセスしてきた閲覧者は、不適切なデータが削除された公開データ記憶部 5 3 0 に格納されたデータのみ入手可能になるので、不適切なデータが一般に公開される

10

20

30

40

50

ことはない。したがって、本実施形態のコンテンツ公開システム500によれば、特定期間内に数多く重複して投稿された、オリジナルコンテンツの存在する、コピーとおぼしき投稿を検出することができ、不適切なコンテンツを見つけるためのベースとなるサンプルコンテンツ(辞書データ)を用意することなく、著作権的に不適切なコンテンツを自動的に検出するという、発明の目的を達成することができる。

【0116】

また、本実施形態のコンテンツ公開システム500によれば、不適切なコンテンツが検出されたとき、アップロードされたコンテンツを削除することにより、不適切なコンテンツが公開されることを防ぐことができる。

【0117】

以上、図面を参照して本発明の実施形態について述べたが、これらは本発明の例示であり、上記以外の様々な構成を採用することもできる。

【0118】

たとえば、上記実施形態のコンテンツ公開システム500は、投稿されたコンテンツを利用者が閲覧できるように公開するコンテンツ公開システムであって、上記実施形態の不適切コンテンツ検出装置(100~300)によって検出された不適切なコンテンツをシステム管理者に提示する提示部(不図示)と、システム管理者が確認した後、システム管理者から削除指示を受け付ける受付部(不図示)と、削除指示に従い、不適切なコンテンツの削除を行う削除部(不図示)と、を有してもよい。

【0119】

また、上記実施形態のコンテンツ公開システム500は、投稿されたコンテンツを利用者が閲覧できるように公開するコンテンツ公開システムであって、上記実施形態の不適切コンテンツ検出装置(100~300)によって相互に類似していると判定されたコンテンツ数が、所定数より多いか否かを判定する判定部(不図示)と、相互に類似していると判定されたコンテンツ数が所定数より多い場合に、当該コンテンツへの利用者のアクセスを自動的に停止する制御部(不図示)と、を有することもできる。

【実施例1】

【0120】

次に、具体的な実施例を用いて本発明を実施するための最良の形態の動作の例を説明する。

図9は、本発明の不適切コンテンツ検出装置の利用のおおまかなイメージを模式的に示した図である。図9において、放送局610は、番組1の放送を行う。ユーザA、B、C、D、Eはそれぞれテレビ放送を端末装置620で受信して、番組1を視聴する(ステップS101、S111、S121、S131、S141)。ユーザA、B、Eは番組を録画して、一部を切り出し、動画V1、V2、V5をそれぞれ投稿サイトにアップロードしたとする(ステップS103、S113、S143)。ユーザC、Dは利用者所有のハンディカム等で撮影した独自コンテンツ(動画V3、V4)をそれぞれ投稿サイトに投稿したとする(ステップS123、S133)。投稿された動画はそれぞれネットワーク602を介して確認サーバ630が受け付ける。確認サーバ630で投稿された動画はそれぞれ不正判定が行われ、適切なコンテンツのみが公開サーバ640にて公開されることとなる。

【0121】

図9で、動画V1、動画V2、動画V5は同一のオリジナルコンテンツの一部を切り出したものであり、その中で同一フレームを含む場合、類似映像として検知可能である。一方、動画V3および動画V4はまったくの独自コンテンツであるため、他のどの動画とも類似しない(図中、動画V3と他の動画との非類似関係を点線の矢印で示し、動画V4の非類似関係については図示していない。)。この結果、図10のようになり、相互類似が確認された、動画V1、V2、V5(図中、実線の矢印で類似関係を示す。)は、不正コンテンツ候補として判定される。

【0122】

10

20

30

40

50

図 1 1 は代表的な類似のパターンをいくつか示している。

図 1 1 ではオリジナル動画 V o の特定の部分を切り出している。図 1 1 中横軸は、時間軸を示しており、動画 V a、V b、V c、V d はオリジナルコンテンツの部分フレームを切り出した形（クリップ）になっている。

【 0 1 2 3 】

たとえば、動画 V a の V a 1 と動画 V b の V b 1 は同一フレームを含み、動画 V b の V b 2 と動画 V c の V c 1 は同一フレームを含み、動画 V c の V c 2 と動画 V d の V d 1 は同一フレームを含む。このように同一フレームを含むため、動画 V a と V b、V b と V c、V c と V d は相互に類似しているといえる。動画 V a と V d は同一フレームが含まれておらず相互に類似しているとはいえないが、動画 V a、V b、V c、V d の相互の接続関係から、動画 V a と V b と V c と V d で相互に類似していると判定可能である。

10

【 0 1 2 4 】

図 1 2 は静止画の場合の例である。たとえば、オリジナル書籍 7 0 2 などをスキャナ等で取り込んだ後に投稿した場合（ステップ S 2 0 2 ）、図 1 2 のように多少は異なるが相互に類似したコンテンツ（画像 7 1 0、7 2 0、7 3 0）が多数投稿される。なお、一致する画像もかなり含まれる。たとえば、画像 7 1 0 はオリジナル書籍 7 0 2 の画像 7 1 2 が傾いて含まれている。また、画像 7 2 0 はオリジナル書籍 7 0 2 の画像 7 2 2 の周囲にのりしろ 7 2 4 が含まれている。さらに、画像 7 3 0 は、オリジナル書籍 7 0 2 の画像の一部が切り取られた画像 7 3 2 が含まれている。

【 0 1 2 5 】

20

これらのような画像も、特許文献 2、特許文献 3 を初めとする各類似度照合エンジンにより、相互の類似コンテンツとして検出することは可能であるため、このような不正コンテンツを削除することが可能になる。音楽データほかでも同様である。

【 0 1 2 6 】

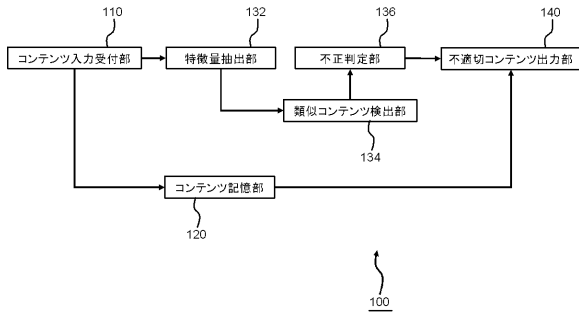
この出願は、2 0 0 7 年 1 0 月 1 9 日に出願された日本出願特願 2 0 0 7 - 2 7 2 9 6 8 号を基礎とする優先権を主張し、その開示の全てをここに取り込む。

【 0 1 2 7 】

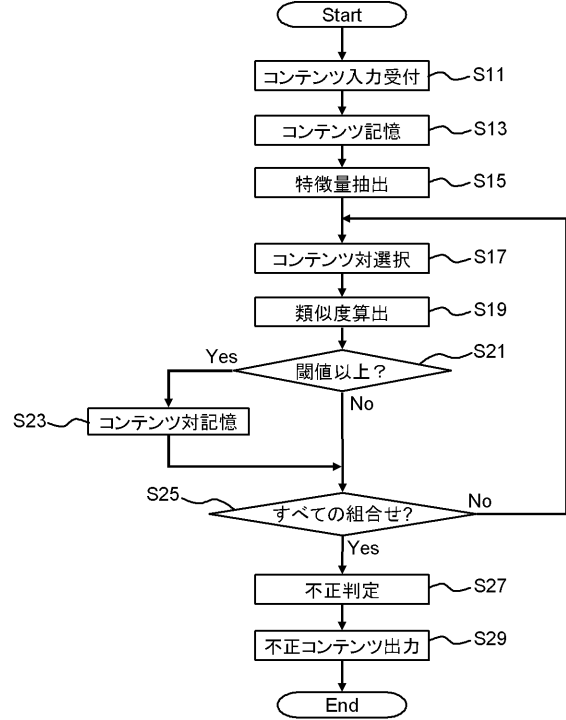
以上、実施形態および実施例を参照して本願発明を説明したが、本願発明は上記実施形態および実施例に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解し得る様々な変更をすることができる。

30

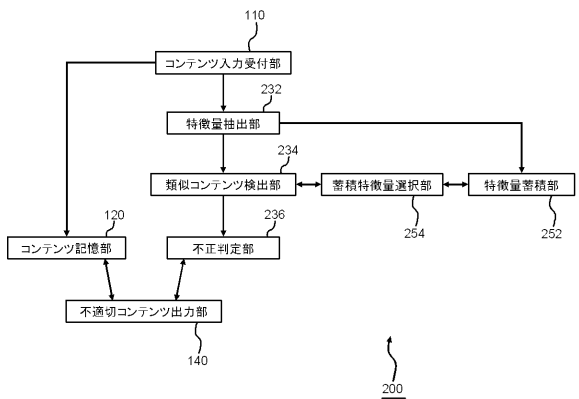
【図1】



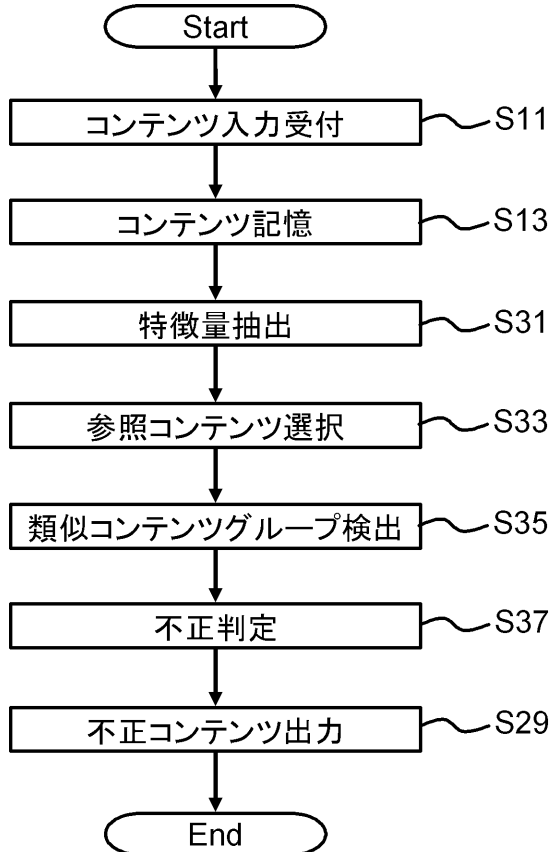
【図2】



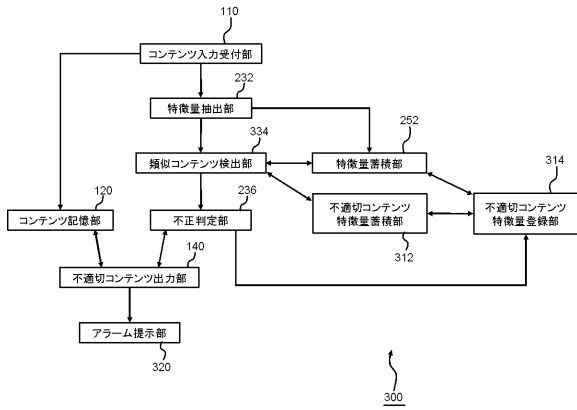
【図3】



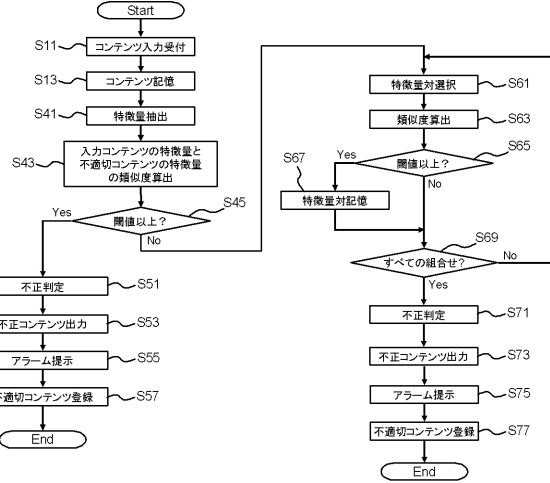
【図4】



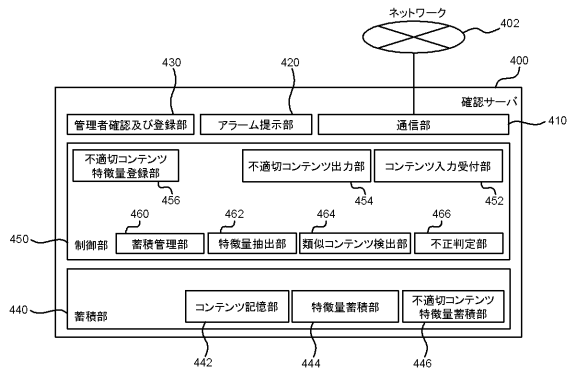
【図5】



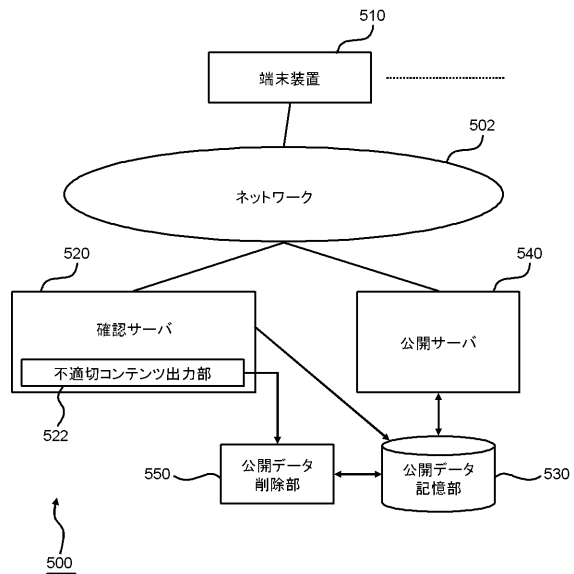
【図6】



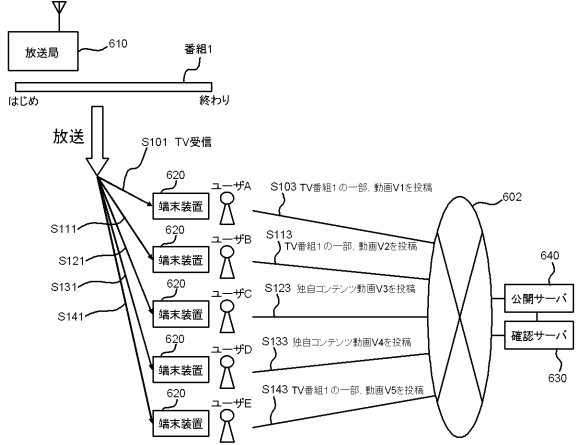
【図7】



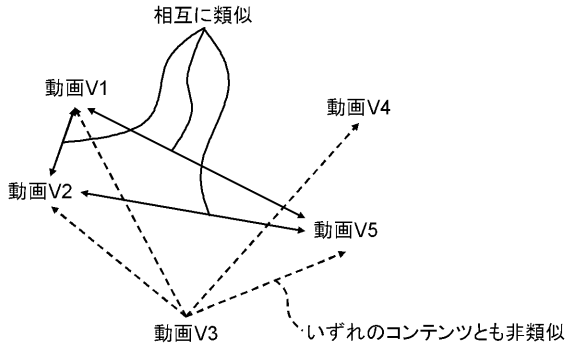
【図8】



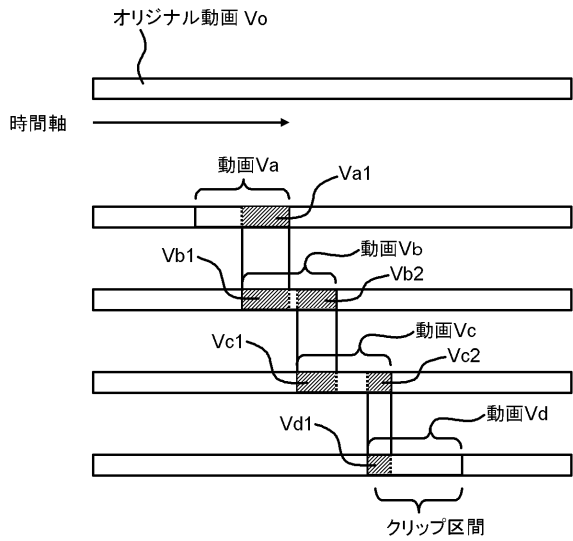
【図9】



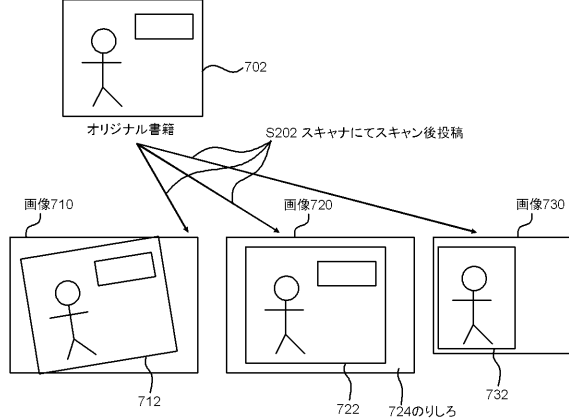
【図10】



【図11】



【図12】



---

フロントページの続き

- (56)参考文献 特開2006-293455(JP,A)  
特開2003-242281(JP,A)  
特開2000-339474(JP,A)  
特開平08-180176(JP,A)  
特開2000-259832(JP,A)  
特開2007-096608(JP,A)  
鈴木 勝也 Katsuya Suzuki, 楽曲協同創作支援システム, 情報処理学会研究報告 Vol. 2007 No. 102 IPSJ SIG Technical Reports, 日本, 社団法人情報処理学会 Information Processing Society, 2007年10月11日, 第2007巻, p.13-17., p.14右欄「2.3 盗作判定」

(58)調査した分野(Int.Cl., DB名)

G06F 21/24

G06F 17/30