総合的な特許ライティング支援環境

IRD国際特許事務所　所長・弁理士　博士（情報学）
特許版・産業日本語委員会委員
谷川　英和
htanigawa@ird-pat.com

目次

1.特許ツールMAP

2.特許ライティング支援環境

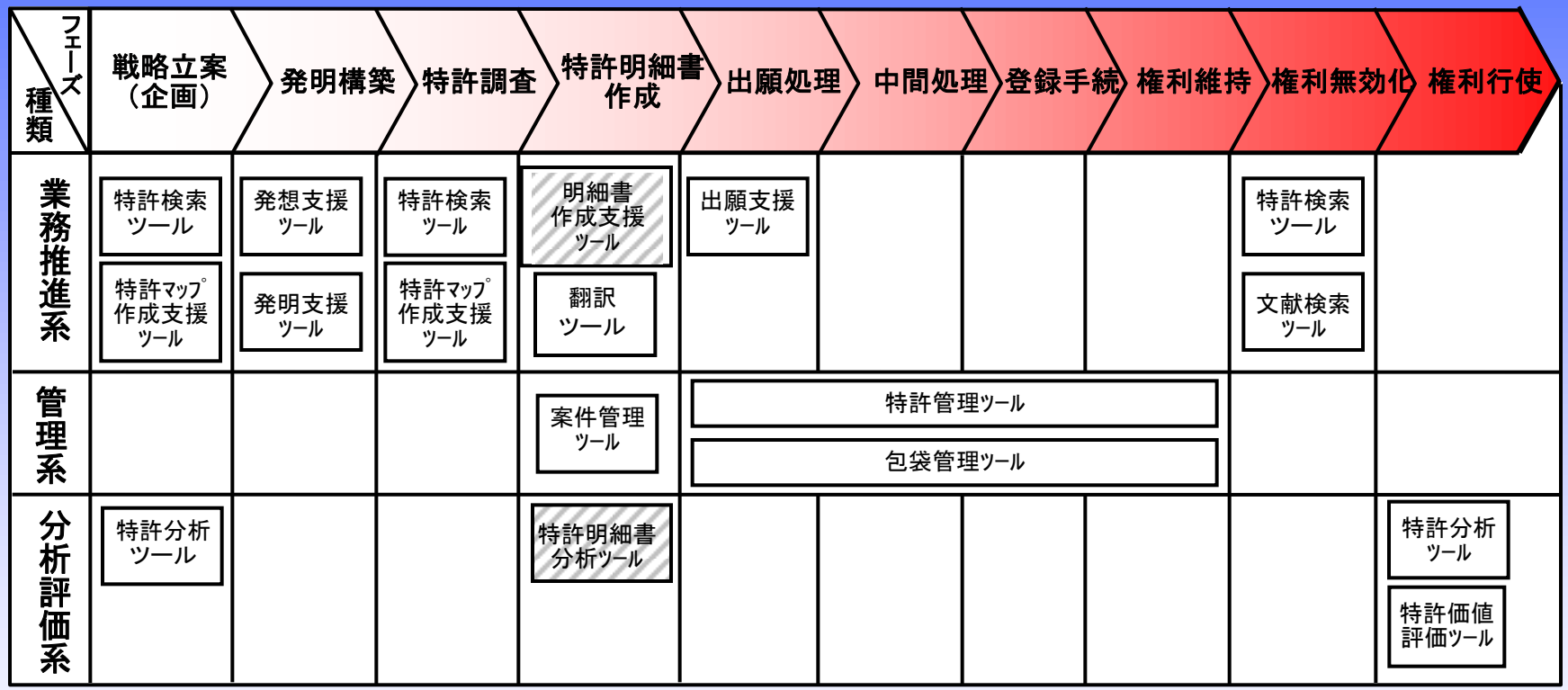
3.特許明細書半自動生成システム（PatentGenerator）

4.読解支援システム （Patent Structure Viewer）

5.可読性診断システム（KnowledgeMeister文書診断 ）

6. 評価結果と今後

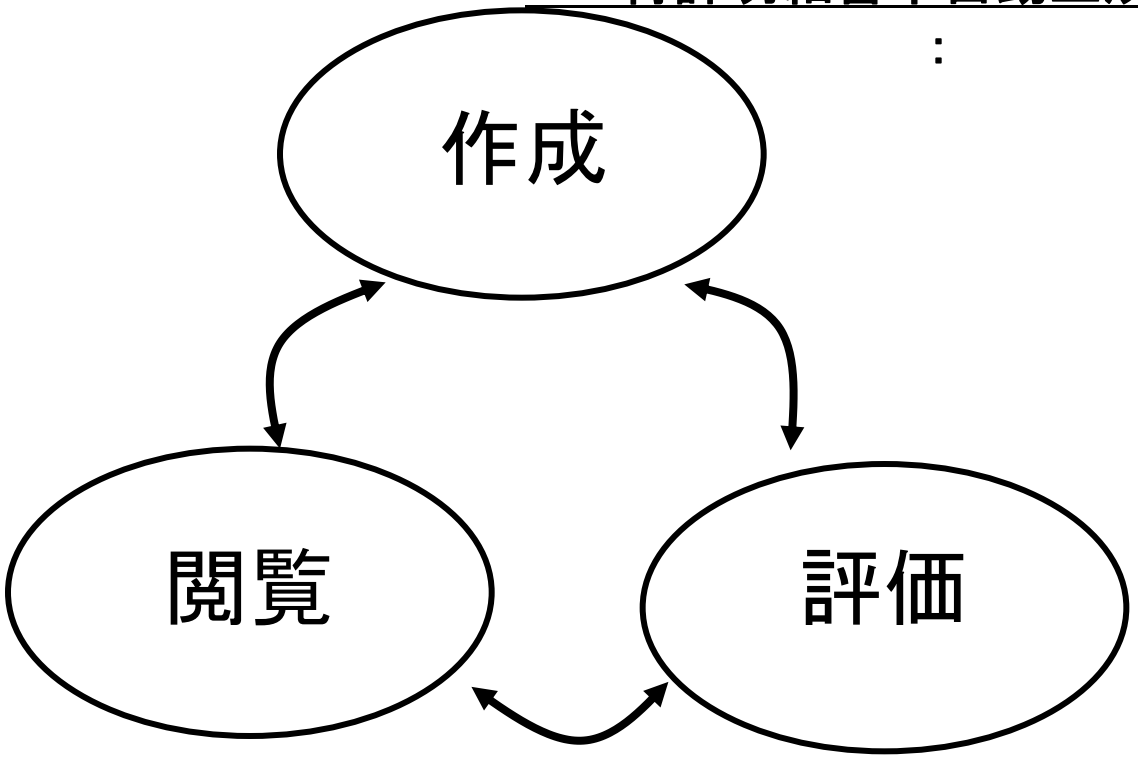
1.特許ツールMAP


フェーズ
種類
戦略立案（企画）
発明構築
特許調査
特許明細書作成
出願処理
中間処理
登録手続
権利維持
権利無効化
権利行使
業務推進系
特許検索ツール
特許マップ作成支援ツール
発想支援ツール
発明支援ツール
特許検索ツール
特許マップ作成支援ツール
明細書作成支援ツール
翻訳ツール
出願支援ツール
特許検索ツール
文献検索ツール
管理系
案件管理ツール
特許管理ツール
包袋管理ツール
分析評価系
特許分析ツール
特許明細書分析ツール
特許分析ツール
特許価値評価ツール

2.特許ライティング支援環境

IRD:特許明細書半自動生成システム

:



インテック:読解支援システム

:

東芝:可読性診断システム

:

特許ライティング支援システムの評価委員

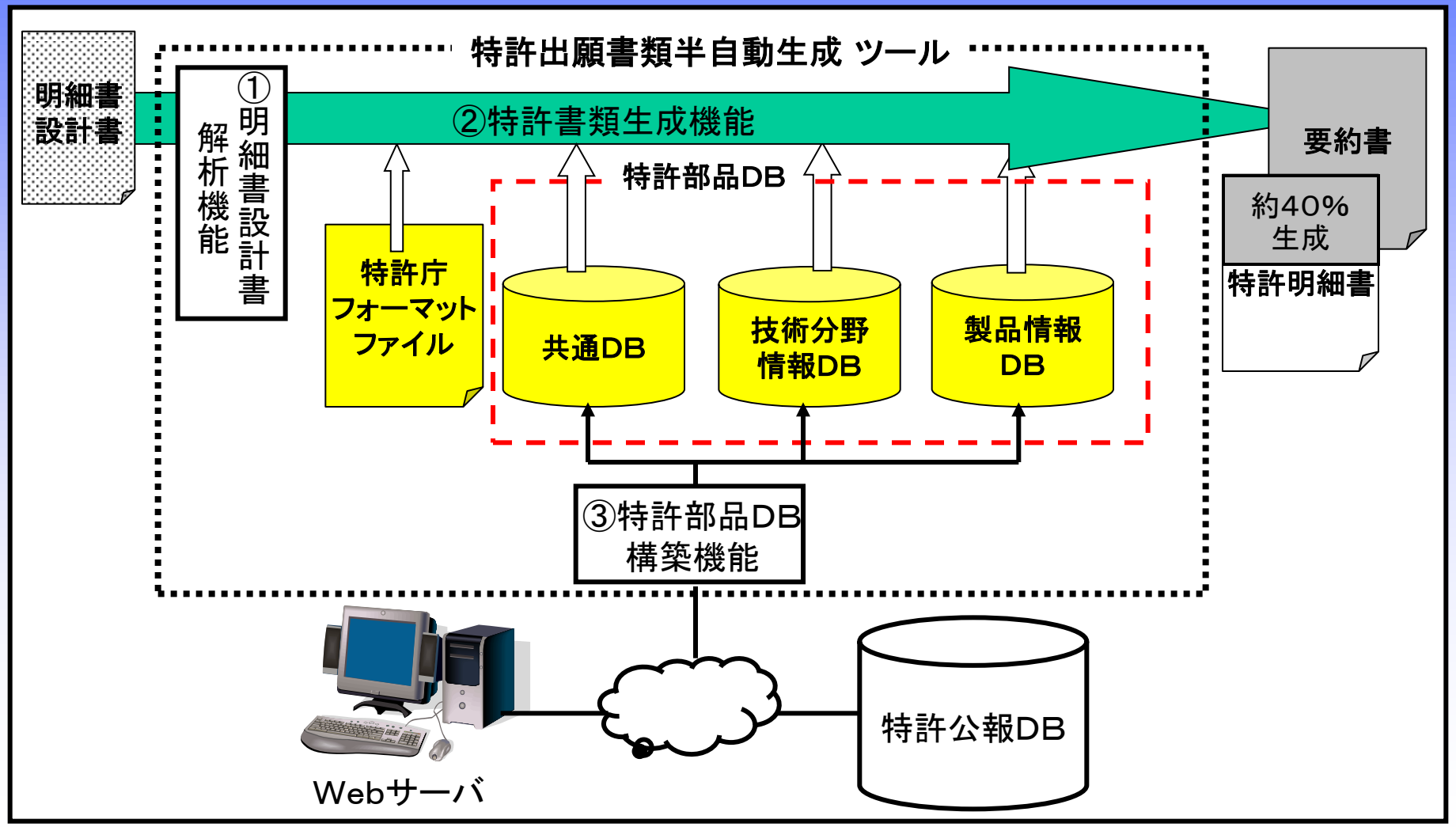
岩永　勇二	平田国際特許事務所　副所長　弁理士
熊野　　明	東芝ソリューション株式会社 プラットフォームソリューション事業部　ソフトウェア開発部　参事
新森　昭宏	株式会社インテック　先端技術研究所 研究開発部　部長
谷川　英和	IRD国際特許事務所　所長　弁理士
松本　征二	知崇国際特許事務所　所長　弁理士
的場　成夫	有限会社　夢屋　代表取締役　弁理士
横山　淳一	富士通株式会社　知的財産権本部 特許統括部　シニアマネージャー　弁理士

3.特許明細書半自動生成システム（PatentGenerator）

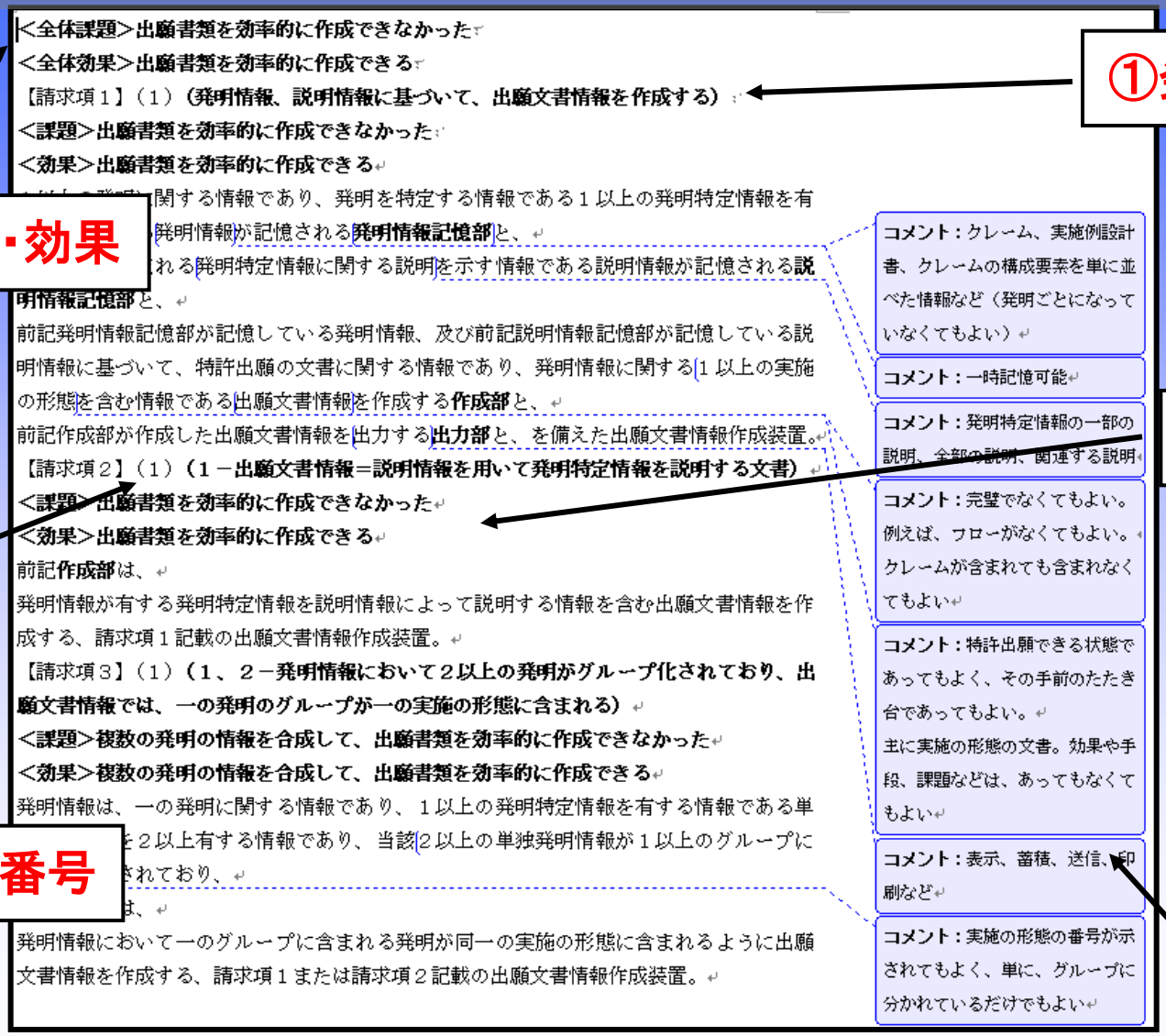
(1)特許明細書半自動生成システム(PatentGenerator)

弁理士のノウハウ、発明を構成する部品の文章をDB化し、利用する

→ 特許書類作成効率1．8倍、特許書類品質1．4倍に！



(2)入力ファイル(明細書設計書)の例



<全体課題>出願書類を効率的に作成できなかった
<全体効果>出願書類を効率的に作成できる
【請求項１】（１）（発明情報、説明情報に基づいて、出願文書情報を作成する）
<課題>出願書類を効率的に作成できなかった
<効果>出願書類を効率的に作成できる
関する情報であり、発明を特定する情報である１以上の発明特定情報を有
発明情報が記憶される**発明情報記憶部**と、
れる発明特定情報に関する説明を示す情報である説明情報が記憶される**説明情報記憶部**と、
前記発明情報記憶部が記憶している発明情報、及び前記説明情報記憶部が記憶している説明情報に基づいて、特許出願の文書に関する情報であり、発明情報に関する１以上の実施の形態を含む情報である出願文書情報を作成する**作成部**と、
前記作成部が作成した出願文書情報を出力する**出力部**と、を備えた出願文書情報作成装置。
【請求項２】（１）（１－出願文書情報＝説明情報を用いて発明特定情報を説明する文書）
<課題>出願書類を効率的に作成できなかった
<効果>出願書類を効率的に作成できる
前記**作成部**は、
発明情報が有する発明特定情報を説明情報によって説明する情報を含む出願文書情報を作成する、請求項１記載の出願文書情報作成装置。
【請求項３】（１）（１、２－発明情報において２以上の発明がグループ化されており、出願文書情報では、一の発明のグループが一の実施の形態に含まれる）
<課題>複数の発明の情報を合成して、出願書類を効率的に作成できなかった
<効果>複数の発明の情報を合成して、出願書類を効率的に作成できる
発明情報は、一の発明に関する情報であり、１以上の発明特定情報を有する情報である単
を２以上有する情報であり、当該２以上の単独発明情報が１以上のグループに
されており、
は、
発明情報において一のグループに含まれる発明が同一の実施の形態に含まれるように出願文書情報を作成する、請求項１または請求項２記載の出願文書情報作成装置。

コメント：クレーム、実施例設計書、クレームの構成要素を単に並べた情報など（発明ごとになっていなくてもよい）
コメント：一時記憶可能
コメント：発明特定情報の一部の説明、全部の説明、関連する説明
コメント：完璧でなくてもよい。例えば、フローがなくてもよい。クレームが含まれても含まれなくてもよい
コメント：特許出願できる状態であってもよく、その手前のたたき台であってもよい。
主に実施の形態の文書。効果や手段、課題などは、あってもなくてもよい
コメント：表示、蓄積、送信、印刷など
コメント：実施の形態の番号が示されてもよく、単に、グループに分かれているだけでもよい

①発明の概要
②実施形態番号
③用語説明
④全体課題・効果
⑤課題・効果

(3)特許庁フォーマットファイル(ひな形情報)

特許明細書の書式
・＜○○＞は、システム変数であり、入力ファイルから抽出して、代入される
・＜＜○○＞＞は、ユーザ定義変数であり、共通DBから検索し、代入される

【書類名】明細書
【発明の名称】＜すべての発明の名称＞
【技術分野】
＜技術分野＞
【背景技術】
＜背景技術＞
【先行技術文献】
　【特許文献】
＜＜特許文献例＞＞
　【非特許文献】
＜＜非特許文献＞＞
【発明の概要】
　【発明が解決しようとする課題】
　しかしながら、従来の＜発明の名称＞においては、＜全体課題＞という課題があった。
　【課題を解決するための手段】
　本第＃の発明の＜発明の名称＞は、＜請求項＞である。
　かかる構成により、＜効果＞。
　【発明の効果】
　本発明による＜発明の名称＞によれば、＜全体効果＞ことができる。

(4)製品情報

製品(構成要素)の部品に関する情報
⇒発明を構成する**構成要素の説明文**
⇒**権利範囲を拡大**するための**説明文**
⇒**特許要件を満たす**ための**説明文**

名称	説明
デコーダ	＜構成要素＞は、デコードされるべき各ソース単語シーケンスについて与えられる補間重みのベクトルに従って、全てのサブシステムからの全てのモデルを線形補間可能である。こうするために、検索に先立って、デコーダはまず、各サブシステムからの句（フレーズ）テーブルをマージしなければならない。全ての句テーブルの句の全てが、デコードの間に用いられる。１つのサブシステムのテーブルで発現するが他のサブシステムのテーブルでは発現しない句も用いられるが、トレーニング中にこの句を獲得しなかったサブシステムによるサポートはない（ゼロ確率）。探索処理は、典型的な多段句ベースデコーダにおけるのと同様に行われる。
翻訳仮説推定手段	＜構成要素＞は、デコーダからのｎベスト出力をとり、ソース及びターゲット分類器に従ったクラスが一致するリストの中で最も高い翻訳仮説を選択する。
統計的機械翻訳装置	トレーニングモジュールによってトレーニングされたモデルと、トレーニングモジュールによって推定された重みWとに基づいて、統計的に翻訳を行う。
格納	＜構成要素＞は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。
＜他＞	＜構成要素＞は、通常、ＭＰＵやメモリ等から実現され得る。＜構成要素＞の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはＲＯＭ等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

(5)技術分野情報

技術分野に関する情報
⇒技術分野に固有の**専門用語の説明文**
⇒**権利強化を図る**ための**定型文**
⇒**文献**としての役割を果たすための**定型文**

名称	説明
ターゲット言語	ターゲット言語は、通常、アラビア語（ａｒ）、デンマーク語（ｄａ）、ドイツ語（ｄｅ）、英語（ｅｎ）、スペイン語（ｅｓ）、フランス語（ｆｒ）、インドネシア語（マレー語）（ｉｄ）、イタリア語（ｉｔ）、日本語（ｊａ）、韓国語（ｋｏ）、マレーシア語（マレー語）（ｍｓ）、オランダ語（ｎｌ）、ポルトガル語（ｐｔ）、ロシア語（ｒｕ）、タイ語（ｔｈ）、ベトナム語（ｖｉ）、中国語（ｚｈ）である。
受け付け	受け付けとは、タッチパネルや、キーボードなどの入力デバイスから入力された情報の受け付け、有線もしくは無線の通信回線を介して送信された情報の受信、光ディスクや磁気ディスク、半導体メモリなどの記録媒体から読み出された情報の受け付けなどを含む概念である。
出力	出力とは、ディスプレイへの表示、プロジェクターを用いた投影、プリンタでの印字、音出力、外部の装置への送信、記録媒体への蓄積、他の処理装置や他のプログラムなどへの処理結果の引渡しなどを含む概念である。なお、送信や蓄積、処理結果の引渡しについては、出力対象が最終的にユーザに提示されるものとする。

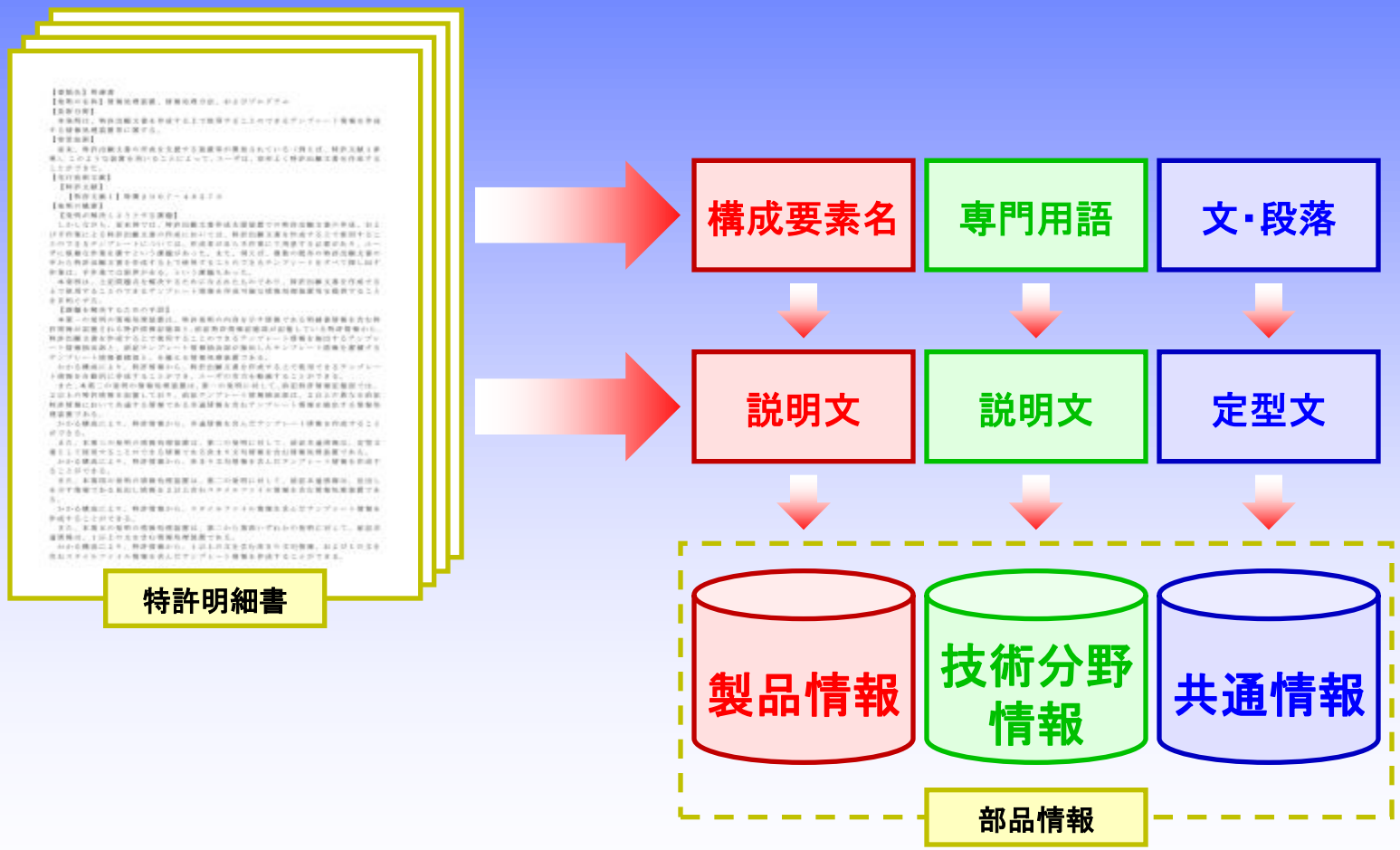
(6)共通情報

特許明細書における**定型文**
⇒**任意の技術分野**において**共通する文章**
⇒**技術分野を問わず共通する文章**
⇒特許明細書における定型文

名称	説明
システム構成1	図１はこの実施の形態のＳＭＴシステム３０の全体構造を示す。図１を参照して、ＳＭＴシステム３０は、クラス依存ＳＭＴモデル、ソース文を分類するために用いられる分類器モデル、及びＳＭＴデコーダ内で用いられる句テーブルをトレーニングするためのトレーニングモジュール４４を含む。トレーニングセット４２はトレーニングデータとして用いられる。トレーニングモジュール４４はさらに、一般ＳＭＴモデルに割当てられる重みＷ１を推定する。重みは、開発セット４０に基づいて推定される。バイリンガルコーパスはクラスに分けられ、さらに、各クラスについてトレーニングセットと開発セットとに細分される。
システム構成2	図２は図１のＳＭＴトレーニングモジュール７４とクラス依存ＳＭＴモデル１１２の３つの集合とを示す詳細なブロック図である。
実施の形態の効果	以上、本実施の形態によれば、～できる。
発明を実施するための形態フッタ	今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

(7)特許部品DB作成システム

特許明細書、特許公報などの**特許文書**から、
特許部品DBを**作成**する(**部品情報**(ひな形情報を除く)を**抽出**する)。



より効率的な部品情報の抽出・蓄積が可能に!!

4.読解支援システム（ Patent Structure Viewer Viewer）

(1) 研究開発の動機

- 「特許請求の範囲」の低可読性の問題
  - 特許請求項の読みにくさ
    - 文長が長く、構造が複雑
      - 主張する権利範囲を1文で記述
      - 読みやすさよりも、正確性・厳密性が優先
      - 専門家コミュニティが形成され、一定の記述スタイルが確立
  - 請求項の引用関係の複雑さ
    - 多数項引用、多重多数項引用
    - 外的付加と内的付加

(2) Patent Structure Viewer について

以下の機能を提供するソフトウエアです

1. 特許請求項の細粒度解析

- 構成要素列挙形式
  - 「～と、～と、～とを備えた○○」
- 順次列挙形式（書き流し形式）
  - 「～し、～し、～する○○」
- 変形構成要素列挙形式
  - 「～と、～と、～とを備え、～は～し、～は～する○○」
- ジェプソン形式
  - 「～において、～を特徴とする○○」

2. 注釈付きクレームツリーの生成

- クレームツリーの自動生成と「電気回路図」的表現での表示
- 内的付加・外的付加の判定
  - 外的付加＝「現状の構成要素の組み合わせに対して別の構成要素を付け加えるもの」
    - 「請求項○に記載の○○であって、～○○を更に備えた○○。」
  - 内的付加＝「上位概念に内包される下位概念を案出するものであり、それには限定、選択が含まれる」
    - 「請求項○に記載の○○であって、前記○○は○○である○○。」
- 構成要素の重ね合わせ

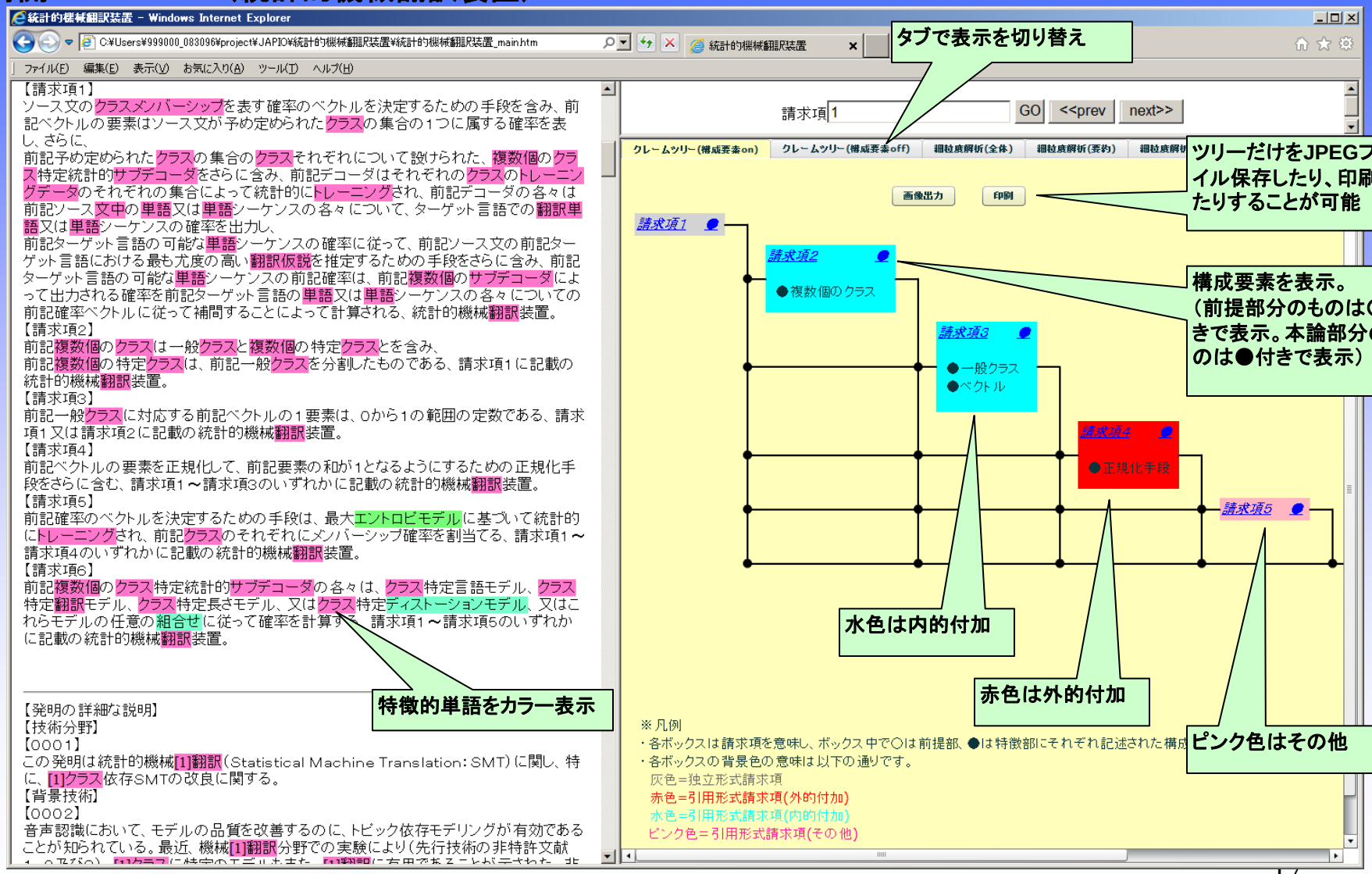
3. 特許請求項における特徴的単語の強調表示（オプション）

- 他の特許書類にあまり出現しない単語を検出し、カラー表示

(3) 実行例①

特開2009-294747（統計的機械翻訳装置）

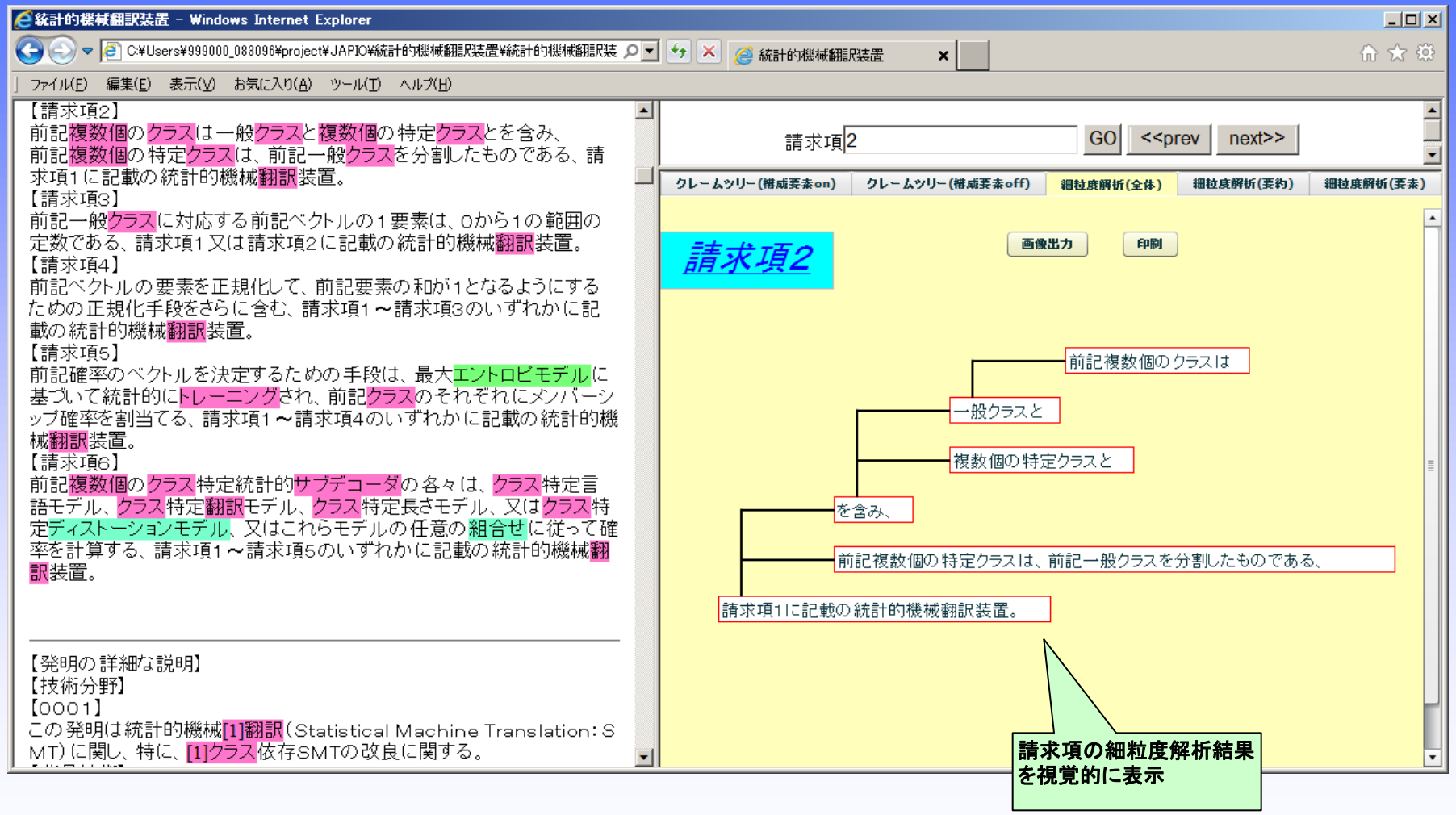
注釈付クレームツリーの表示、特徴的単語のカラー表示



(4) 実行例②

特開2009-294747（統計的機械翻訳装置）

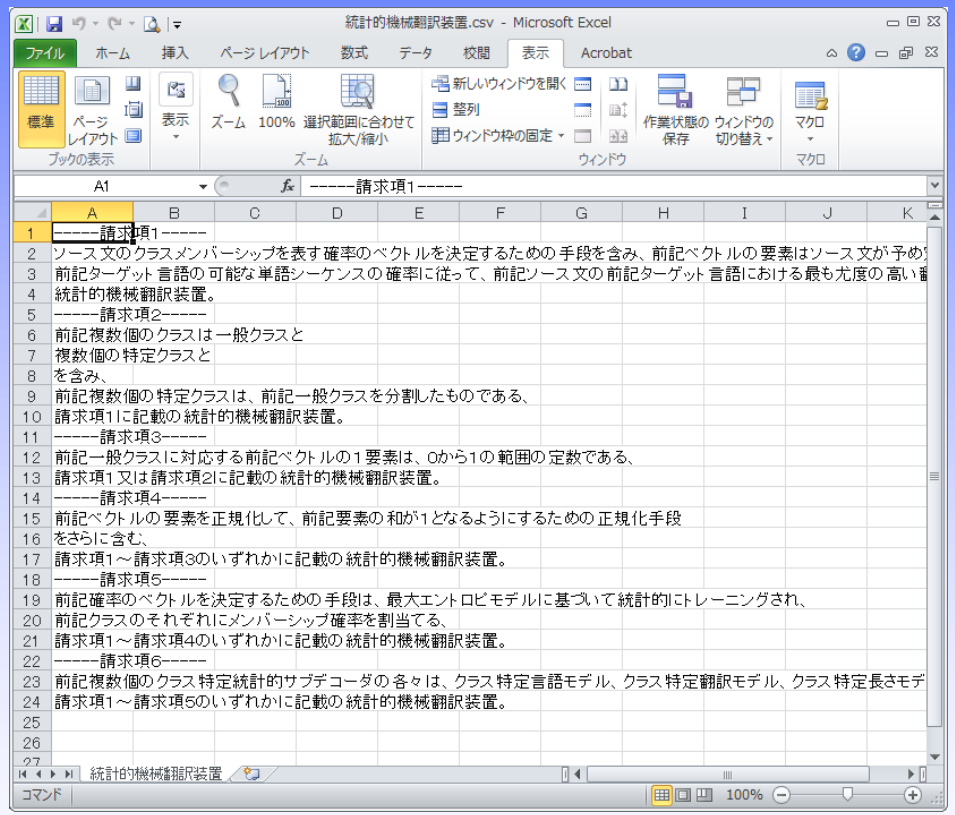
特許請求項の細粒度解析結果表示



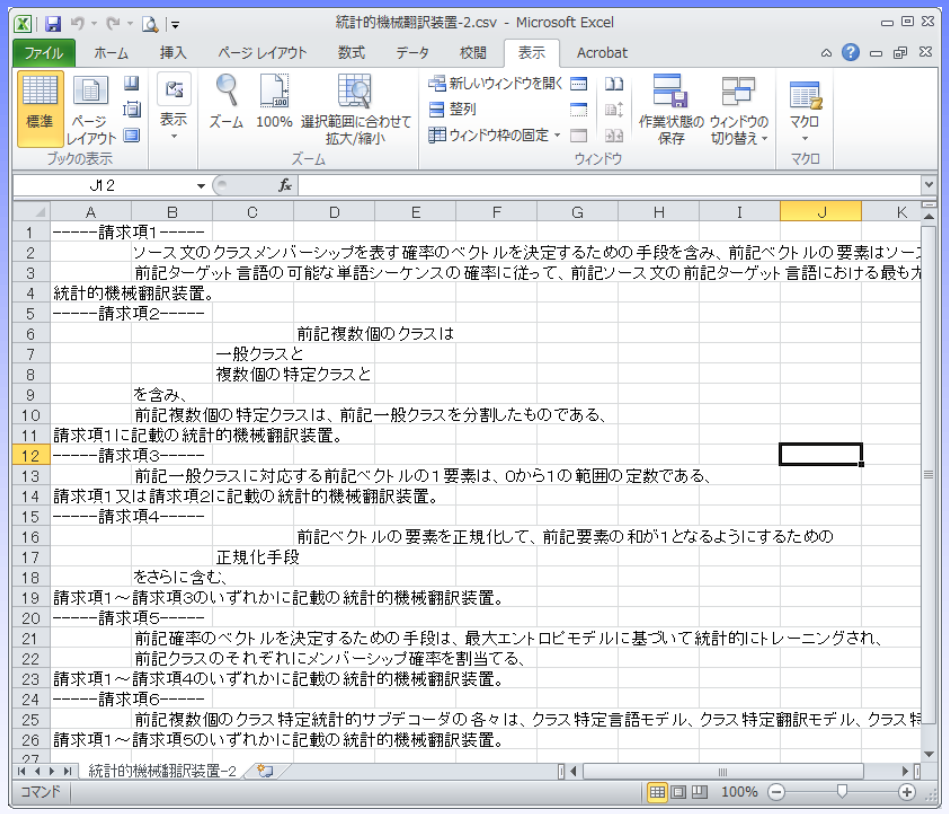
(5) 実行例③

特許請求項の解析結果のCSV出力

■ 粗い解析結果と細粒度解析結果の2種類のCSVを自動出力



粗い解析結果



細粒度解析結果

5.可読性診断システム（KnowledgeMeister文書診断 ）

(1)業務文書の品質

業務文書の品質

内容の品質
※文書を設計するとき決まる。例えば、記載の中身、詳細度など。

表現の品質
※文書を執筆するとき決まる。

表現の正確性
※記載した表現と記載しようとした表現が異なる。
例えば、誤字脱字、かな漢字変換の誤りなど。

表現の可読性
※記載した表現が人に理解しやすいかどうか、コンピュータで処理しやすいかどうか。

表現の整合性
※記載した情報に整合性があるかどうか。
例えば、引用先と引用元の不一致、数値の不整合、表記揺れなど。

(2)機械翻訳への活用例

原文

「顧客用文書をカラー、社内用文書を白黒で印刷する」

診断なし

診断あり

診断結果：
「カラー」の直後に「で印刷し」が省略されているようです。

可読性診断システム

修正後：
「顧客用文書をカラーで印刷し、社内用文書を白黒で印刷する」

診断結果に基づく修正

機械翻訳システム

英訳結果

Color and the document for in-house use are printed for the document for customers in black and white.

×

中国語訳結果

在黒白印刷顧客用文书彩色、公司内部用文书。

×

英訳結果

The document for customers is printed in color and the document for in-house use is printed in black and white.

○

中国語訳結果

以彩色印刷顾客用文书，以黑白印刷公司内部用文书。

○

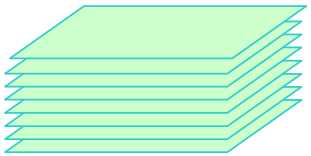
Copyright 2012 Toshiba Solutions Corporation

(3)可読性診断技術の活用シーン

- **機械翻訳の前処理**
  - 曖昧性のある表現が指摘される
  - 原文を修正し、翻訳の精度を向上
- **翻訳コストの低減**
  - 分かりやすい翻訳原文が作成できる
  - 翻訳者からの問合せ、翻訳者の誤解による誤訳が減少
- **人が理解しやすい文章の作成**
  - 仕様書などの技術文書で、正しい情報伝達を支援
  - 理解度が増し、業務効率が向上

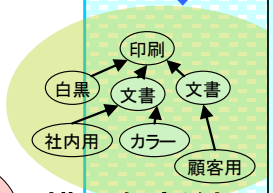
(4)可読性診断技術の構成

翻訳対象文書



日本語構文解析エンジン

可読性診断システム



構文解析結果

規則	内容
曖昧な係り受け	係り受け先の候補が複数あり、どれが正しいのか判断しにくい
複合語	辞書に登録されていない複合語に訳しにくい接辞が含まれている
述語の省略	連用中止となる述語が省略された並列構造
…	

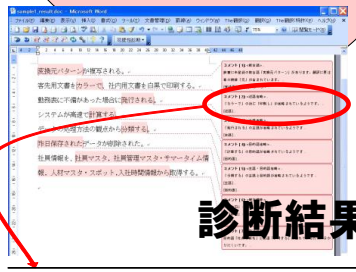
診断規則

診断モジュール

- 曖昧な係り受けの診断
- 複合語の診断
- 述語の省略の診断
- 助詞「は」の診断
- ・・・

診断結果に対する確認と修正

ユーザ



診断結果

コメント[2]：＜述語省略＞

「カラー」の後に「で印刷し」が省略されているようです。

[述語] で印刷し

付加情報

翻訳者

機械翻訳

(5)技術的特徴

- 翻訳しにくい文の特徴を考慮した診断規則に基づいて、文章の可読性を診断
- 日本語構文解析の結果を使用し、文の構造から誤訳しやすい箇所や曖昧な箇所を検出
- 翻訳の精度を向上するための付加情報や選択肢を生成

コメント[2]：＜述語省略＞
「カラー」の後に「で印刷し」が省略されているようです。
[述語] で印刷し

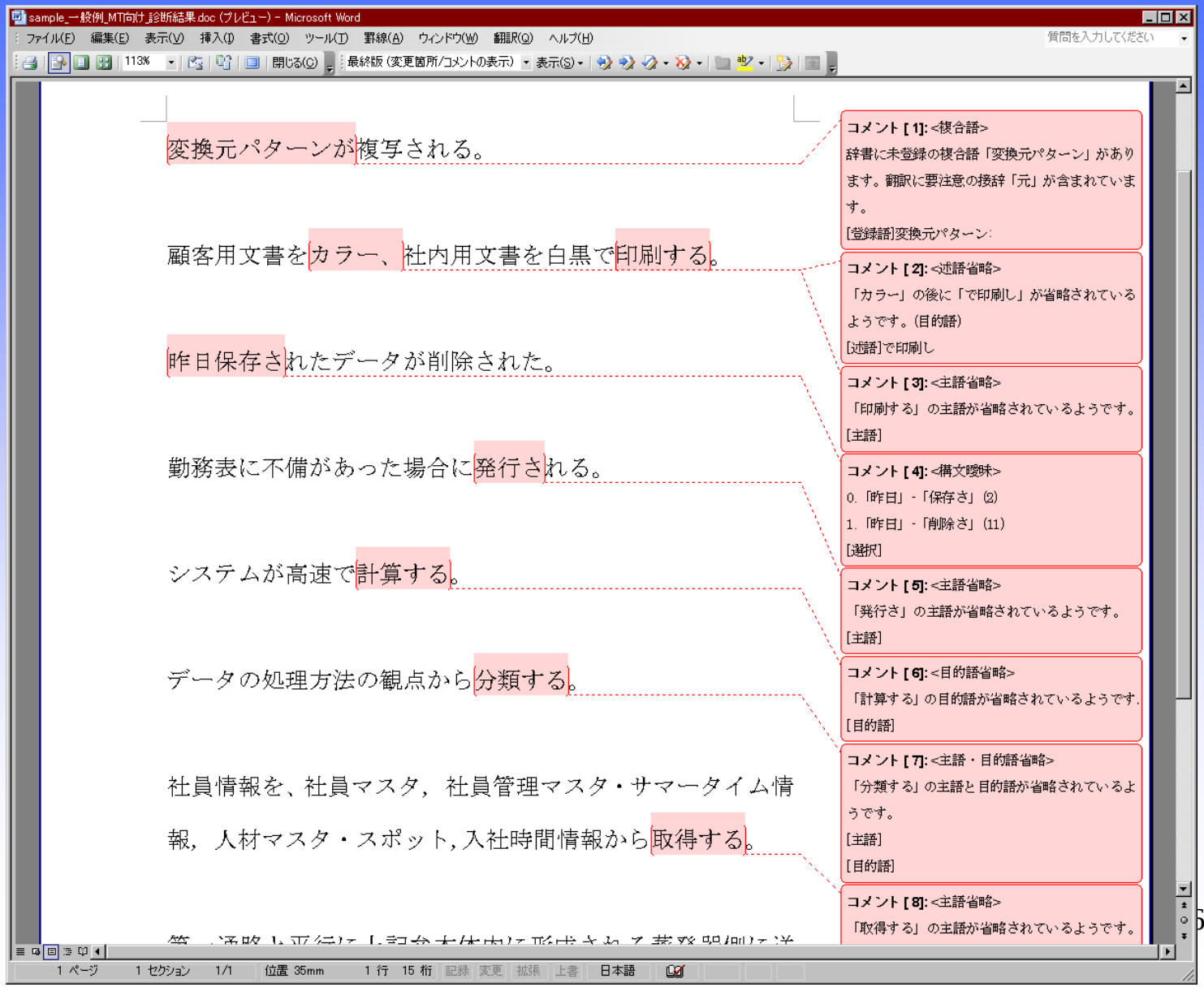
付加情報

昨日保存されたデータが削除された

コメント[3]：＜曖昧な係り受け＞
「昨日」－「保存」
「昨日」－「削除」

選択肢

(6)可読性診断結果の出力例



(その他)産業日本語との関連

(1)**特許明細書半自動生成システム(PatentGenerator)**

・産業日本語ライティングルールに従った特許部品DBを準備すれば、ルールに従った明細書の出力が可能 (明細書の標準化が可能)

(2)読解支援システム (Patent Structure Viewer)

・産業日本語ライティングルールに従った場合、解析精度を一段と向上できる可能性がある

・今後の開発により、産業日本語ライティングルールに従った構造を有するか否かのチェックが可能となる可能性がある

(3)可読性診断技術(KnowledgeMeister文書診断)

・可読性診断の判定基準と産業日本語ライティングルールとの共通性

・ライティングの支援システムや支援環境としての活用

6.評価結果と今後

1.特許ライティング支援システムを構成する各ツールの評価

→ 概ね良好（使用すれば効果がある）

2.各ツールの改善案のまとめ、改善

→ PatentGenerator: 化学分野等において、関連特許を自動検索し、動的に特許部品DBを作成する等の機能が必要

KnowledgeMeister: 特許書類に特化した用語等のチェックが必要

3.特許ライティング支援環境のあるべき姿の提案

→ さらに「教育支援環境」が必要

ご清聴ありがとうございました