

民事判決のオープンデータ化検討PT（第7回）

令和3年1月26日（火）

13:00～14:00

弁護士会館17階1703会議室

議 事 次 第

- 1 冒頭挨拶等
- 2 実証実験の報告
- 3 今後の論点整理について
- 4 意見交換
- 5 次回以降の日程について

（配布資料）

- 1 民事判決のオープンデータ化検討PT構成員名簿及び出席者一覧（第7回）
- 2 株式会社 Legalscape 提出資料

参考1 民事判決のオープンデータ化検討PT第8回以降の日程（案）

民事判決のオープンデータ化検討PT 構成員名簿及び出席者一覧（第7回）

2021年（令和3年）1月26日

◎：座長 ○：座長代理

	構成員	出席者（第7回・1/26）
◎	菊地裕太郎（日本弁護士連合会前会長）	菊地裕太郎（日本弁護士連合会前会長）
○	早稲田祐美子（日弁連法務研究財団専務理事）	早稲田祐美子（日弁連法務研究財団専務理事）
※	日本弁護士連合会	菰田 優（前事務総長）
※	山本和彦（一橋大学教授）	山本和彦（一橋大学教授）
※	町村泰貴（成城大学教授）	町村泰貴（成城大学教授）
※	小塚荘一郎（学習院大学教授）	小塚荘一郎（学習院大学教授）
	一般財団法人司法協会	松本英司（複写事業部長） 川端素子（出版事業部長）
	一般財団法人法曹会	清水 孝（主事）
※	株式会社 Legalscape	八木田樹（代表取締役・最高経営責任者）
※		津金澤佳亨（最高執行責任者）
※		城戸祐亮（取締役・最高技術責任者）
※		久本空海（ソフトウェアエンジニア）
※	ウエストロー・ジャパン株式会社	中村裕哲（執行役員）
※	株式会社有斐閣	高橋 均（常務取締役）
※	第一法規株式会社	田中英弥（代表取締役社長）
※		川原崎晶子（出版編集局編集第一部長）

	オブザーバー	出席者（第7回・1/26）
※	内閣官房	坂本三郎（内閣審議官）
※	法務省	渡邊英夫（法務省司法法制部参事官）【代理】
※		大野晃宏（法務省民事局参事官）
※	最高裁判所事務総局	村田斉志（最高裁判所事務総局総務局長）

	事務局	出席者（第7回・1/26）
	日弁連法務研究財団	大坪和敏（事務局員）
	日本弁護士連合会	藤原靖夫（事務次長）

※ Web 会議システムによる出席者

Confidential

民事判決オープンデータ化検討PT

自動仮名処理に係る実証実験

－追加実験内容及び、最終結果の報告－

2021年1月26日
株式会社Legalscape

- 本実証実験（'20年10月-'21年1月実施）では、民事判決オープンデータ化検討に資することを目的とし、機械による仮名処理の技術的可能性を実験・評価した上で、仮名処理の実務体制構築の方向性オプション・主要検討項目のたたき台を整理した
 - 前回の実験経過報告（'20年11月PT時報告）では、機械処理の2つの段階「①対象語句の特定」「②語句属性の特定」のうち、①における機械処理性能を初期的に確認・報告した
 - 以降本日までに「①の性能向上の取組み」及び「②も含めた機械処理の実装・性能評価」を実施した
- 機械処理性能は、「①対象語句の特定」において再現率・精度ともに**94%前後**（箇所ベース）を達成し、「②語句属性の特定」精度は**98%**（属性グループベース¹⁾）を達成した
 - とくに、①においては5割超の判例で100%の正解（機械によって完全に正しい仮名処理が行われた）となり、既存の日本語新聞や海外判決文等における自然言語処理性能と比較し、高い性能を達成
 - 今後、更なる研究開発等により一定の性能向上は不可能ではないと推察されるものの、他方で機械処理のみによる100%の性能実現は原理的に不可能であるため、実務上は人手による追加修正の要否・程度の検討が不可欠
- 人手による修正方針には複数のオプションがあるが、いずれにおいても費用・手間等は（今後継続的に本PTの目的・条件等に立ち返り確認する必要があるものの）実現に向けた詳細検討が可能な範囲である
 - 人手による確認・修正の方針のオプションには、大きく全文を確認・修正する方針か、機械出力の周辺のみを確認・修正することで、一定の誤りを許容する代わりに確認・修正の手間を削減する方針がある
 - また、いずれの方針においても人手による修正を効率化するツールの開発が可能であり、仮名化に要する時間を現状比で1/6～1/20程度に削減することが可能な想定
 - 今後の実務体制構築にあたっては、仮名処理システムの検討・開発・運用に加え、データ公開の規模や方式等を踏まえた全体の公開システム・実務オペレーション体制に係る検討を要する

1. 語句属性の特定精度の定義については本編p.11に記載



1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
3. 実務体制構築の方向性オプション（議論用叩き台）

背景

民事判決のオープンデータ化を行う上で、仮名処理をどのように／どの程度、施すべきかが検討されている

- 特に、技術的な観点からは、判決の大量公開にあたって、一定の機械的な自動処理が前提とされる中で、機械的な処理の可能性、精度評価が必須
- 同時に、機械による自動処理が完全でないことも想定されるため、人手による追加修正の要否・経済性の評価・検討も必須

V

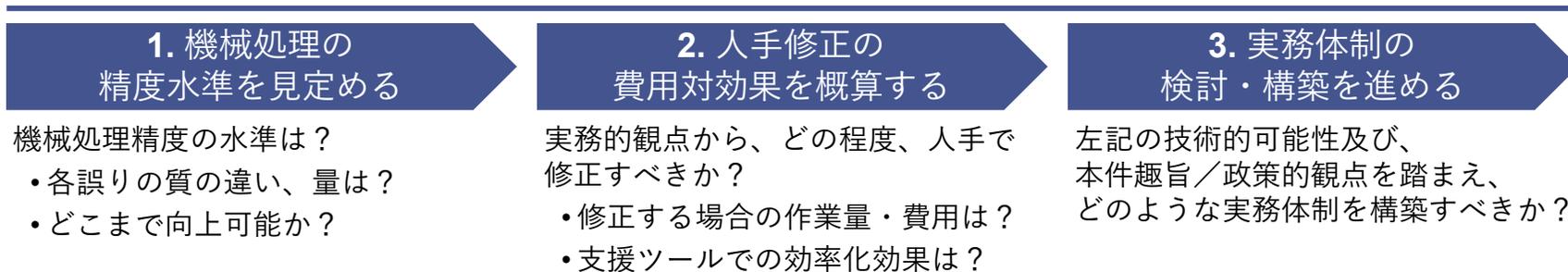
目的

本実験は、機械処理精度の評価、及び、人手による修正の費用対効果を検証・推定することで、取るべき仮名処理手法／体制構築の検討に資することを目的とする

実証実験の流れ・位置付け

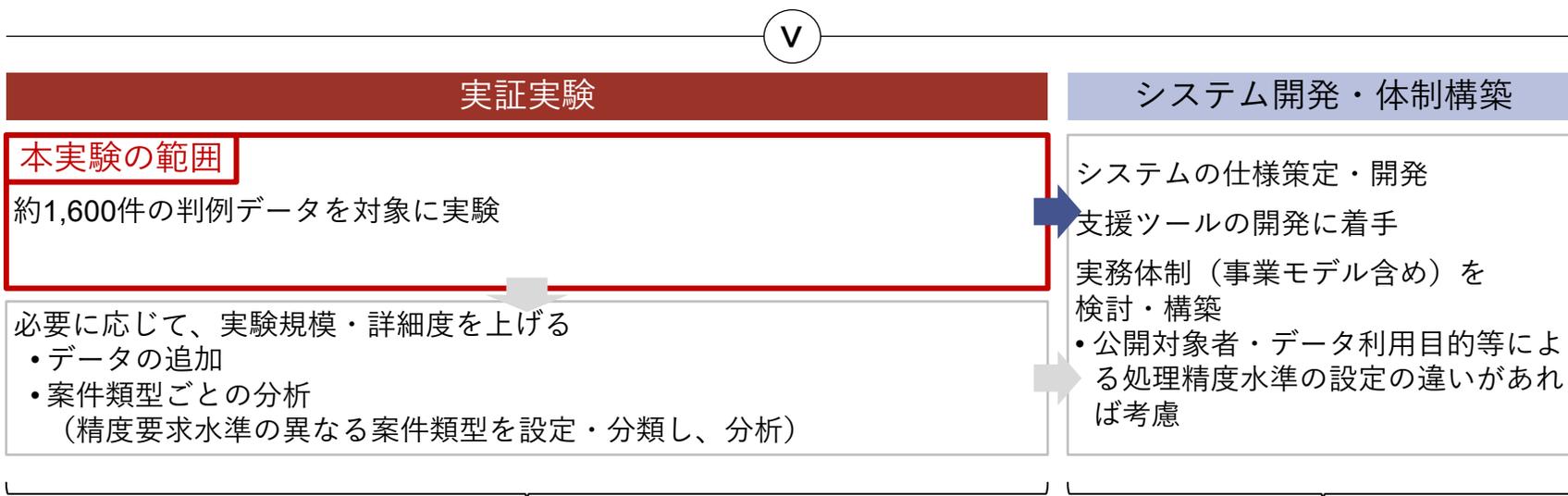
- 約1,600件の判例を対象に実験を行い、自動仮名処理の技術的な可能性を検証・評価
- 技術的可能性の評価を踏まえ、実務体制構築の検討材料を提示

仮名処理の実務体制を構築するまでの流れ



検討論点

検討／実施アプローチ



本資料にて結果を報告

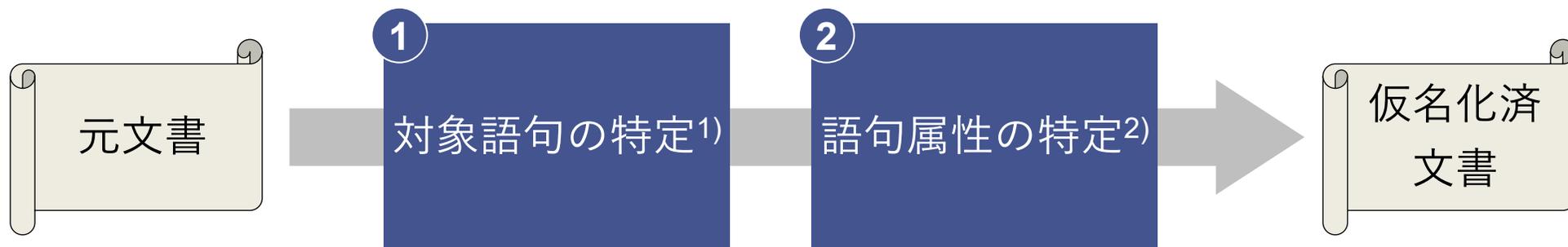
本資料にて検討の枠組・材料を提示



1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
 - ① 実験のアプローチ
 - ② 機械処理の精度水準
 - ③ 人手修正の費用対効果
3. 実務体制構築の方向性オプション（議論用叩き台）

機械処理の流れ

①対象語句の特定、②語句属性の特定の2つの機械処理を施す



原告山田太郎（以下「山田」という）は東京都渋谷区にて...

- 山田太郎
- 山田
- 渋谷区

- 同一: 山田太郎 = 山田
- 住所: 渋谷区

原告X（以下「X」という）は東京都（以下略）にて...

1. 言語処理分野の研究での「固有表現抽出」に該当
2. 言語処理分野の研究での「関係抽出」などに該当

機械処理の流れ：② 語句属性の特定

置換語の選定には、グルーピング・タイプ分類・名付けの各段階における判定を要し、各判定の精度が最終的な処理精度に影響を及ぼす



仮名処理対象の語句全体の中から、同じラベルを付けるべき語句同士をグループ分け

各グループに対して、人名 / 住所 / 車両番号 / 郵便番号、等のタイプを分類

各グループに対して、タイプに則した適切な仮名を付与

処理誤りの意味合い

①②それぞれの機械処理における誤りは、異なる実務上の問題をもたらす

「①対象語句の特定」での誤り

ケースI：
仮名の「漏れ」

山中太郎（旧姓佐藤）は...



X（旧姓佐藤）は...

個人名の漏洩
→ プライバシーリスク

ケースII：
仮名の「過多」

山中の道で...



Xの道で...

文意の不明／変化
→ 閲覧者の利用性
・ 権利の問題

「②語句属性の特定」での誤り

ケースIII：
仮名記号の不一致

山中太郎（旧姓佐藤）は...



X（旧姓Y）は...

文意の変化
→ 閲覧者の利用性
・ 権利の問題

- ・ プライバシーリスクの問題は、①の結果のみが影響
- ・ 閲覧者の利用性／権利上の問題は、①②両方の結果が影響

誤りの種類と評価指標：①対象語句の特定（1/2）

「仮名漏れ（偽陰性）」と「仮名過多（偽陽性）」という2種類の誤りがある

誤りの種類

評価指標

		正解	
		仮名対象	一般語句
機械出力	仮名実施	A 正答	B 仮名過多 (偽陽性)
	仮名せず	C 仮名漏れ (偽陰性)	D 正答

$$\text{精度} = \frac{A}{A + B}$$

仮名実施のうち何割が正解か
(全く特定しなければ100%)

トレードオフ
の関係

$$\text{再現率} = \frac{A}{A + C}$$

仮名対象のうち何割を仮名できたか
(全文字を特定すれば100%)

- プライバシーリスクには「仮名漏れ（偽陰性）」が影響
- 閲覧者の利用性／権利には「仮名過多（偽陽性）」が影響

誤りの種類と評価指標：①対象語句の特定（2/2）

仮名漏れ／仮名過多には、さらに「完全な誤り」と「部分的な誤り」がある

正答

仮名実施（機械出力）

小山田太郎

仮名対象

人手修正には
文書全体の確認が必要

完全な誤り

仮名漏れ

仮名せず（機械出力無し）

小山田太郎

仮名対象

仮名過多

仮名実施（機械出力）

小山田太郎

一般語句

機械出力の確認のみ
による人手修正が可能

部分的な誤り

漏れ + 過多

仮名実施
（機械出力）

小山田太郎

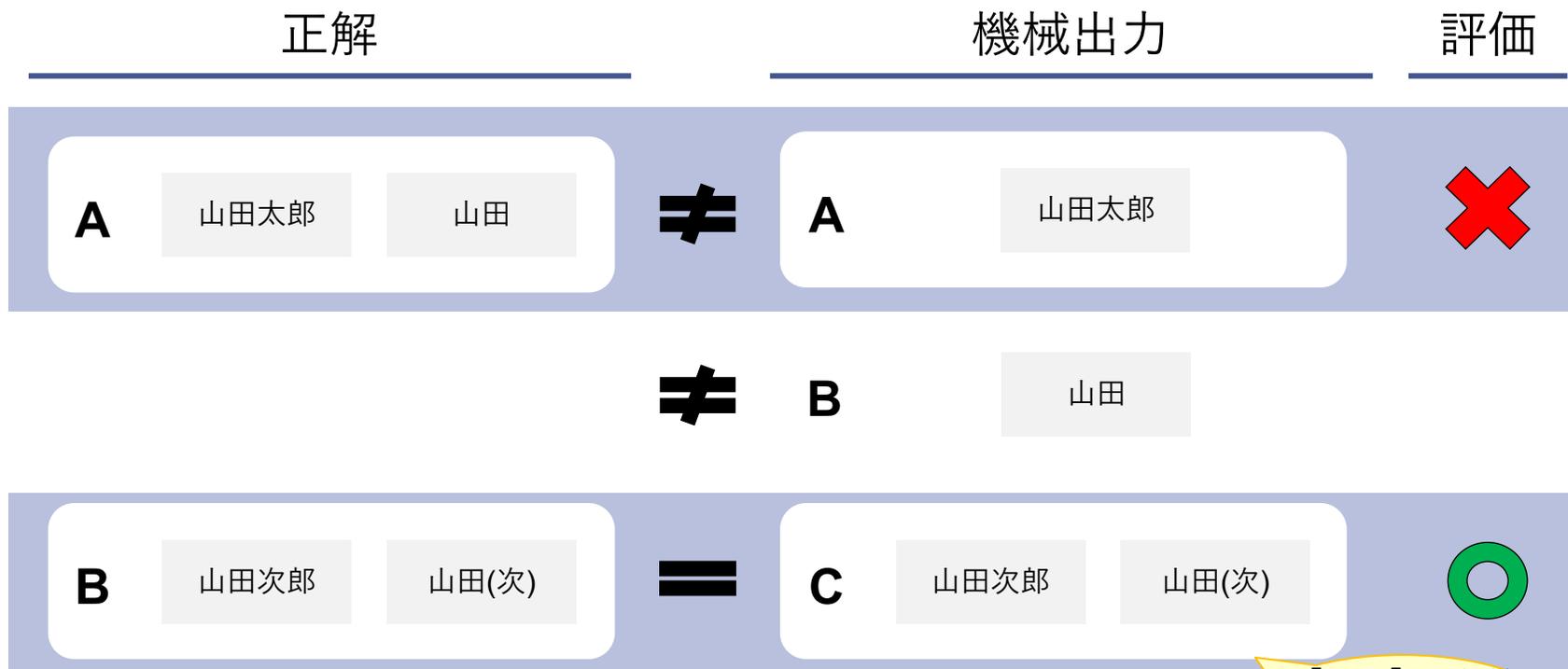
仮名対象

- 人手修正をする場合、修正対象を「完全な誤りのうちの仮名過多」と「部分的な誤り」に絞ることで、修正コストを下げる事が可能

（＝文書全体を確認するのではなく、機械出力のみ確認することで修正の手間を減らす）

誤りの種類と評価指標：② 語句属性の特定

仮名対象語のグループの一致割合を「精度」として定量評価



「B」「C」など
仮名記号自体の
一致は問わない

精度
(グループ一致割合)

$$= \frac{\text{正解と一致する機械出力グループ数}}{\text{正解グループ数}} = \frac{1}{2} = 50\%$$



1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
 - ① 実験のアプローチ
 - ② 機械処理の精度水準
 - ③ 人手修正の費用対効果
3. 実務体制構築の方向性オプション（議論用叩き台）

実証実験の
最終結果

技術詳細
(性能改善の取り組み)

実証実験の最終結果：全体像

- 判決公開に向け、仮名処理実務に機械処理を採用／組込ことは、十分に検討に値する
- 本実験結果は、自然言語処理における1つの研究としても、一定の成果を実現できた

1

対象語句の特定

2

語句属性の特定

再現率（仮名漏れ）

精度（仮名過多）

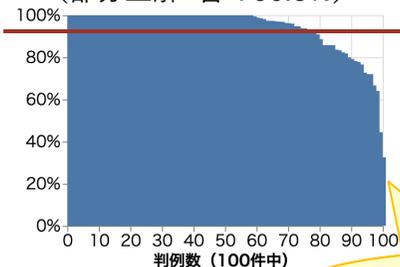
精度（グループ一致割合²⁾）

仮名箇所数ベース

〔 1単語でも
複数回カウント 〕

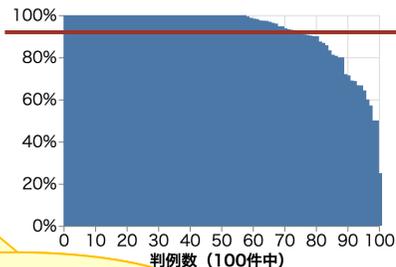
94.5%

（部分正解¹⁾含め96.8%

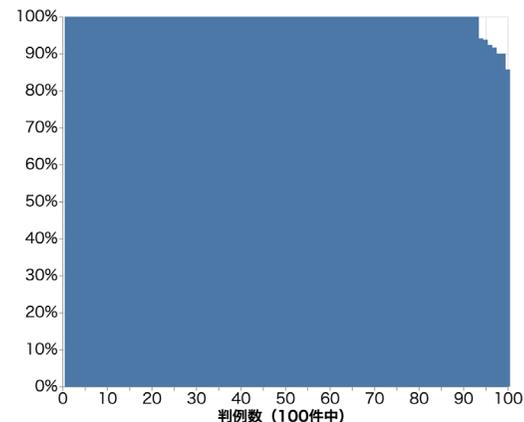


93.4%

（部分正解¹⁾含め96.4%



98.0%



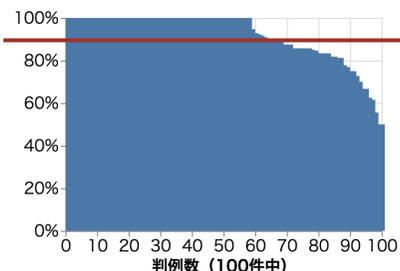
出現回数が多い単語
における誤りが比較的小さい

仮名单語種類
ベース

〔 1つの単語は
1回カウント 〕

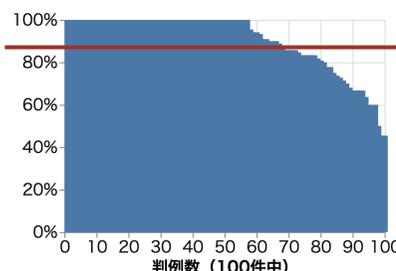
91.8%

（部分正解¹⁾含め96.4%



87.6%

（部分正解¹⁾含め94.5%



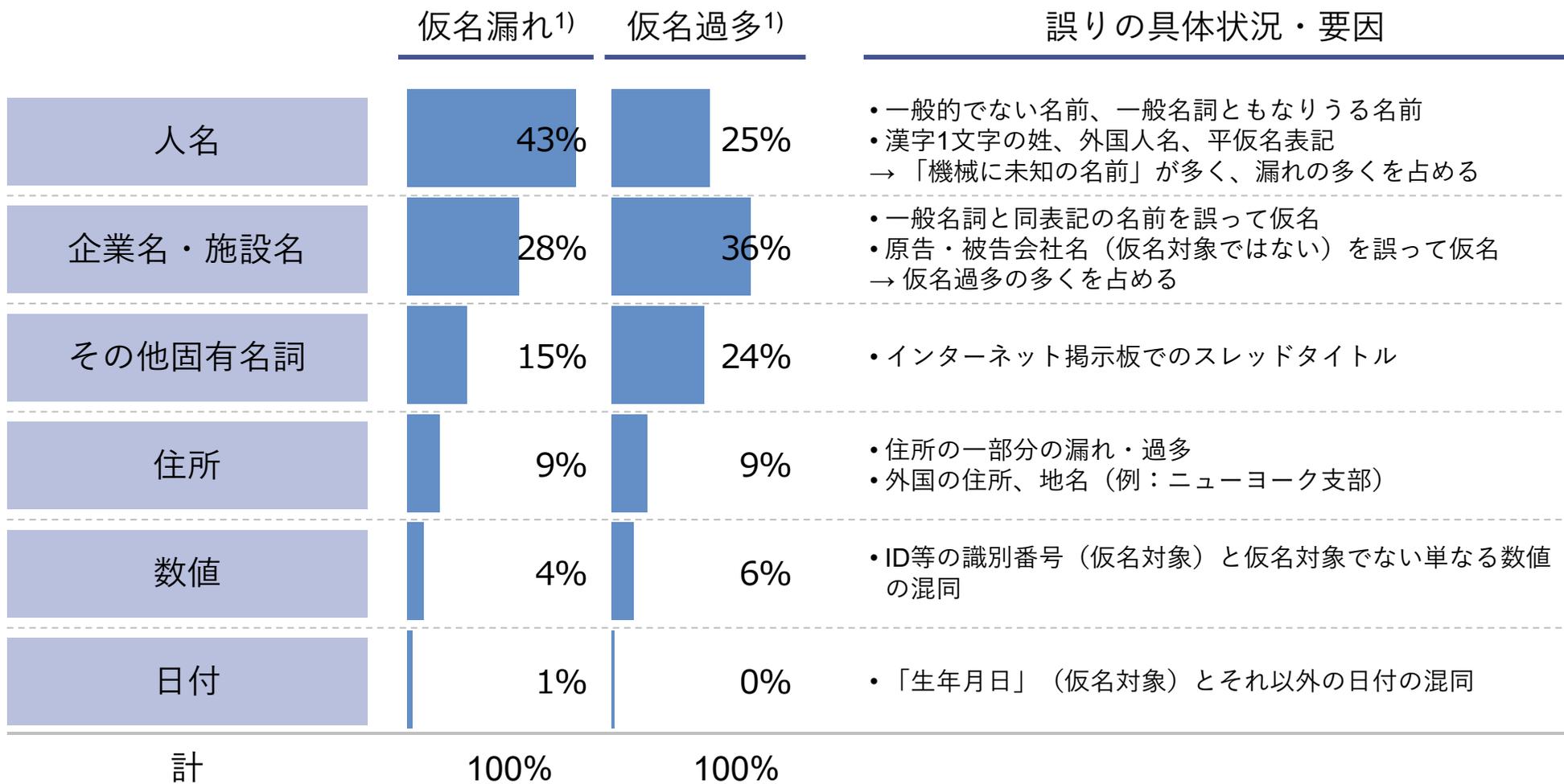
※正解の仮名箇所に対して、
②を実施した際の精度

5割の判例は修正を必要とせず、必要とする場合でも、平均7箇所・3単語の修正に留まる

- 「部分正解含め」た再現率／精度とは、p.10記載の「正答」及び「部分的な誤り」の両方を正解とした場合の再現率／精度を指す
- グループ一致割合の定義は、p.11参照

実証実験の最終結果：詳細：① 対象語句の特定

- 誤り全体を仮名対象の語句タイプ別に分類
- 誤りの80%以上は、人名、企業・施設名、その他固有名詞の認識誤りによる



本実験にて各種精度改善（後頁にて詳述）を施した上でも、なお誤るケースは、「未知の人名・ニックネーム」／「一般名だが固有名」等の機械的判断が難しいケース

実証実験の最終結果：詳細：① 対象語句の特定

- 網羅の困難さから、仮名漏れが発生
- 逆に一般名詞との混同などにより、仮名過多も発生

人名の誤り

仮名漏れ
非定形の名

正解
ネットネームの **tanaka@net**

機械出力
ネットネームのtanaka@net

人名の誤り

仮名過多
一般語と混同

正解
被告は金を要求した

機械出力
被告は**金**を要求した

企業名の誤り

仮名漏れ
既知でない名

正解
被告は、**すみれフラワー**に ...

機械出力
被告は、すみれフラワーに ...

企業名の誤り

仮名過多
一般語と混同

正解
被告は、国立病院に ...

機械出力
被告は、**国立病院**に ...

実証実験の最終結果：詳細：① 対象語句の特定

- 自由度の高い固有名詞や数値は網羅が困難で仮名漏れが発生
- 住所の誤りはほとんどの場合において部分的な仮名過多・漏れ

固有名詞の誤り

仮名漏れ
自由記述名

正解
ネット掲示板「東京グルメ案内」

機械出力
ネット掲示板「東京グルメ案内」

住所の誤り

仮名過多
部分的なミス

正解
東京都北区北部地域

機械出力
東京都北区北部地域

数値の誤り

仮名漏れ
未知の形式

正解
ログインID: 12345

機械出力
ログインID: 12345

日付の誤り

仮名漏れ
生年月日

正解
被告は1月1日に息子Aを出産

機械出力
被告は1月1日に息子Aを出産

ルールベースによるシステムで100%に近い精度を達成

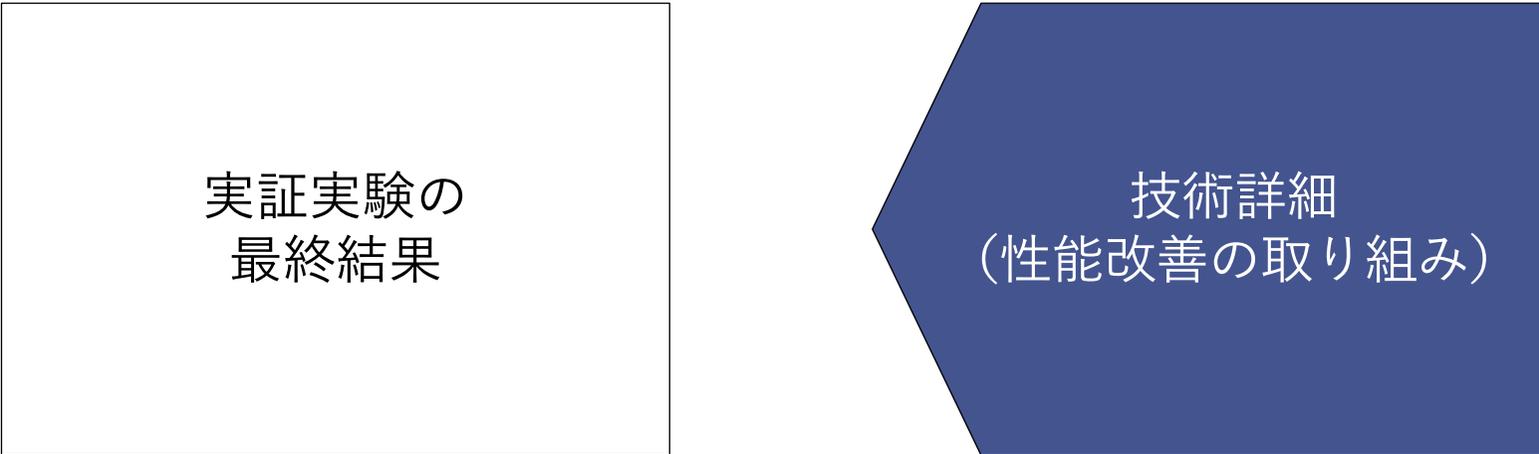
機械的アプローチ

- 古典的なルールベースによる手法のみ
 - チューニングしやすく再現性が高い
- 採用したルール
 - 「○○（以下○○という。）」などの明示的な記述を利用
 - 「フルネーム」と「姓」「名」の対応づけ
 - ひらがな・カタカナ表記の対応づけ（山田＝ヤマダ）

機械処理の結果

- 正解率：98.0 %
 - ※1 「正解の仮名語グループ」（仮名記号）のうち「機械処理で正しく出力したグループ」の数の割合。ステップ①は完璧と仮定し独立に評価
 - ※2 旧姓・通称の区別は今回評価対象外とした
- 誤り分析
 - 明示的でない略称
 - ◆ 「東京タナカ病院」 → 「タナカ病院」
 - 一意に特定できない人名
 - ◆ 「田中太郎」「田中次郎」が登場する判例での「田中」
 - 同表記だが区別される名前
 - ◆ 「田中が所有する田中ビル」 → 「Aが所有するBビル」
 - その他、明示的でない別表記
 - ◆ ニックネーム「田中直樹」 → 「ナオ」
 - ◆ インターネット上の名前「田中」 → 「TANAKA@PC」

- ステップ②の誤りは人間が読む際には容易に発見できるため、利便性・情報公開性の毀損は限定的と思料
- ステップ①の誤りが後続の②に伝播することを踏まえても、①の方がより重要



実証実験の
最終結果

技術詳細
(性能改善の取り組み)

技術詳細：性能改善の取り組み

- 主な改善アプローチを網羅的に検討・採用し、性能改善を達成
- 大規模統計モデルBERT及び個別ルールにより、仮名漏れ的大幅削減に成功（仮名過多も減少）

	採用手法	再現率 (仮名漏れ)	精度 (仮名過多)
前回PT時点	spaCy	85.2%	92.4%
最終結果	BERT+ルール	94.5%	93.4%

各手法での改善幅	再現率	精度
i spaCy+法律辞書	+ 0.3%	(0.6%)
ii BERT	+ 8.8%	+ 0.8%
ii BERT+生データ	+ 7.7%	+ 0.6%
iii BERT+ルール	+ 9.3%	+ 1.0%

一通りの
改善手法
を実施し、

最も改善幅の大きい**BERT+ルール**による
改善手法を最終結果として採用した

更なる改善余地は小さいため、次のステップでは、仮名処理実務としての要求性能／
人手修正方針検討に進むことが望ましいと思料。次頁以降、各改善手法の詳細を解説

i 手法：spaCy + 法律用語辞書の追加

- 法律用語辞書の用語の出現が少なく、性能への影響なし
- 辞書追加による更なる性能向上は、企業名や稀な人名を入れる必要

追加辞書の詳細

- 法律用語辞書として2種類の言語資源を受領
 - 有斐閣様『法律用語辞典』
 - 商事法務様『会社法コンメンタール事項索引データ』
- これらのデータから、既存の機械処理用辞書に未収録の1.2万語を整備し利用
- 今回対象とする判例データの全文約900万語中、0.9%のみが該当する法律用語であった
 - (例) 「弁論の全趣旨」、「不法行為」、「訴訟代理人」

追加結果

- 性能への影響はなし
 - 追加された語彙は専門用語・学術語も多く、判例本文中の出現頻度は小さい上、仮名対象と識別が困難な語がほとんど含まれていない
 - 例えば「訴訟代理人」を「訴訟/代理人」と2単語と解釈していたとしても、結果への影響はなし

→むしろ人名辞書・企業名辞書等のさらなる追加が有効

尚、本実験では、BERTの学習に利用したWikipedia125万記事に含まれる固有名詞をカバーしており、更なる固有名詞の充実のためには、特殊・レアな人名等の辞書が必要

ii 手法：大規模な統計モデル（BERT） + 生データ

判例データ以外の大規模言語データを活用した統計モデル（BERT）の導入により、特に仮名漏れが大幅に改善

大規模統計モデル（BERT）の概要と結果

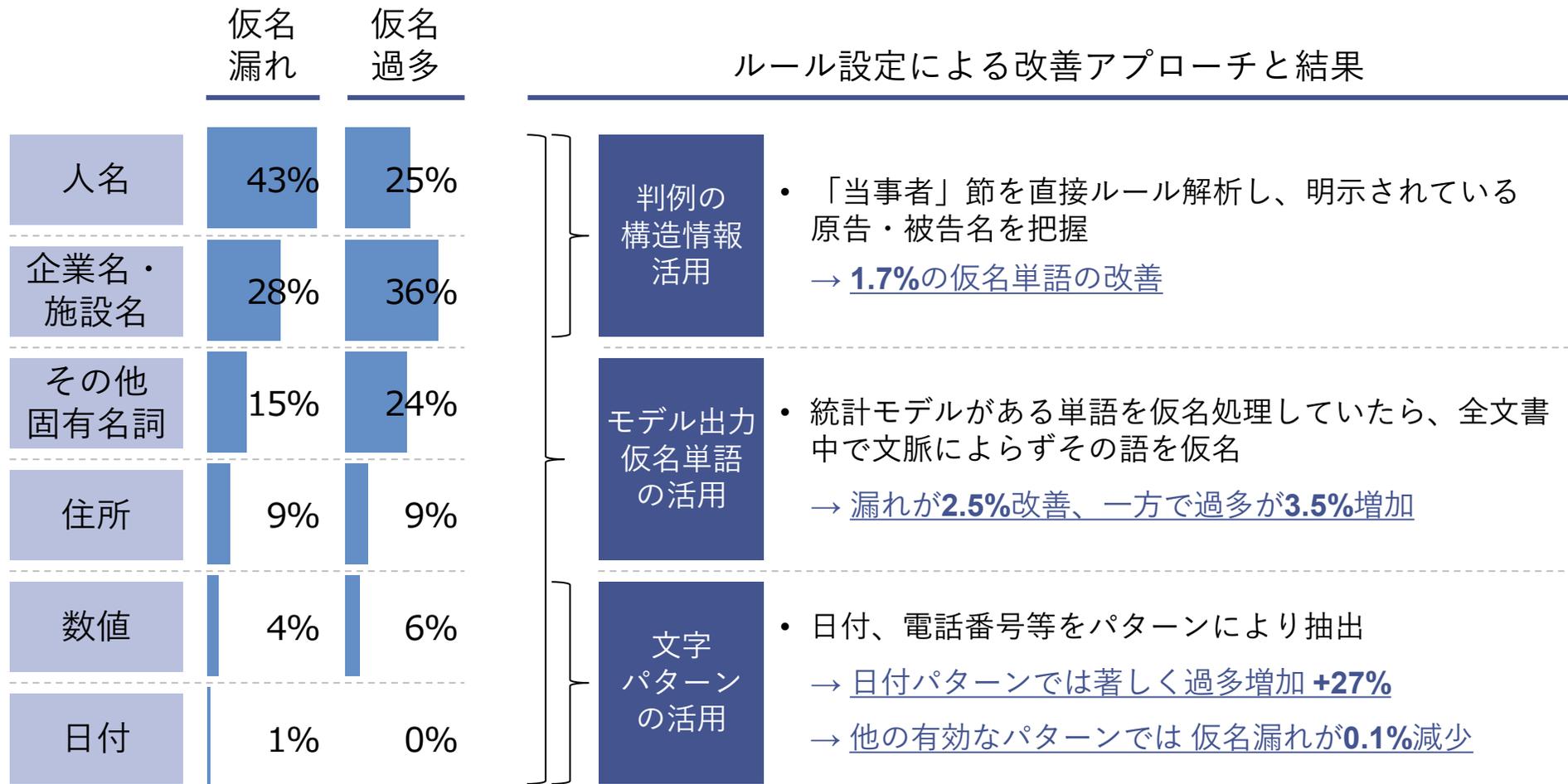
- 大規模なデータにより事前に学習した超大規模の統計モデルを活用
 - Google社による「BERT」
 - 今回は、日本語Wikipedia全記事（125万記事）を利用して学習
- 大規模なため一般的なPCでは利用できず、特殊な計算機環境が必要
- 前回PT時に比べ、特に仮名漏れが大幅に改善
85.2% → 94.0%

生データ（第一法規様判例データ）の活用

- Wikipediaに加えて、第一法規様から受領した2.5万件の仮名処理後判例データを大規模統計モデルの学習に活用
 - 仮名処理後のみであり、処理前との対応はない
 - 約300万行、1.0GBのテキスト
- 大幅な性能向上は確認できなかった
 - 他の研究事例と比較して小規模（参考: Google社 - 33億語のテキスト）
 - 仮名処理後のデータであり固有名詞情報が存在しないため有効性が小さいと想定

iii 手法：大規模な統計モデル（BERT）＋ルールの追加

- 一部分のみ、ルールにより誤り修正が可能
- ほとんどの誤りは固有名詞であり、ルール対応は困難

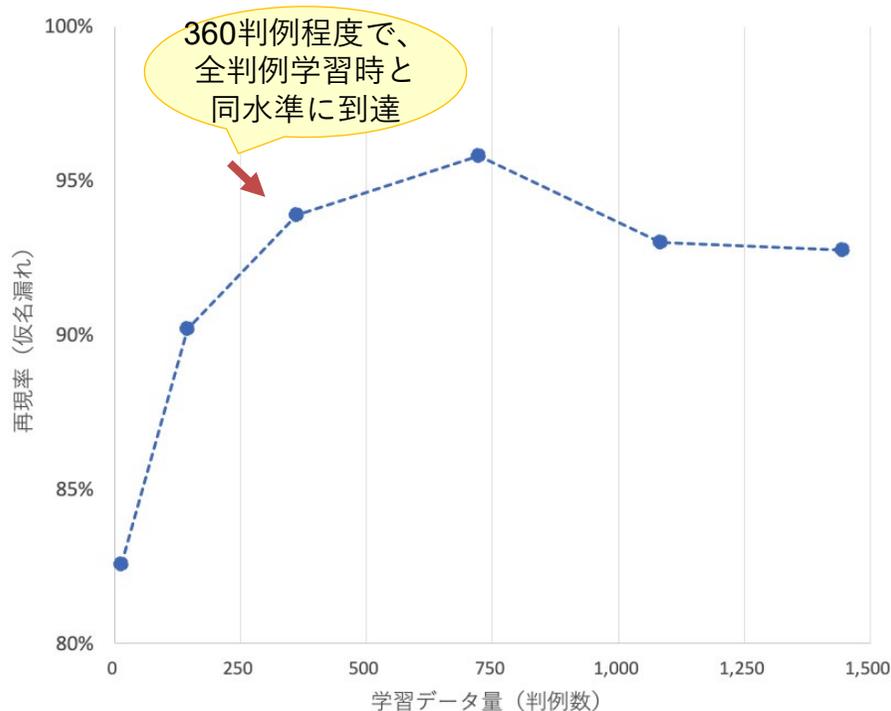


上記を総合しての最大の改善量
漏れ: 仮名单語 2.0% (仮名箇所 0.5%)
過多: 仮名单語 0.3% (仮名箇所 0.2%)

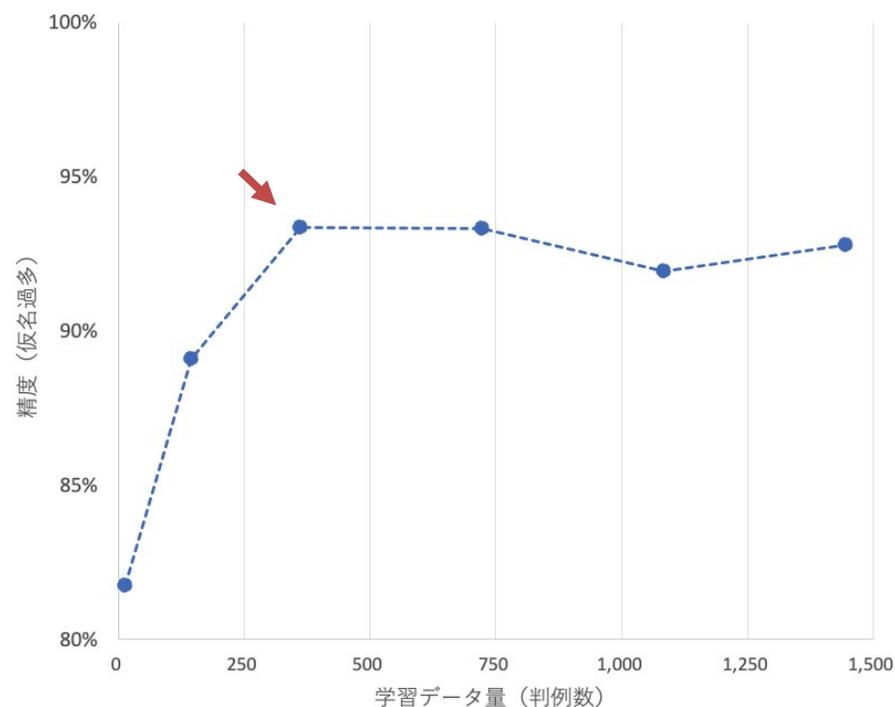
参考) 学習データの追加による性能改善余地

現時点で入手可能性の比較的高い範囲の判例数（最大でも数万程度想定）の範囲においては、更なる判例データの追加による性能向上は限定的な可能性が高いと史料

再現率（仮名漏れ）の学習データ量別変化



精度（仮名過多）の学習データ量別変化



360判例（現在利用できるデータの**25%**）でも全判例利用による学習と同程度の性能（追加するデータの中身によっては、逆に性能が低下する可能性も有り）

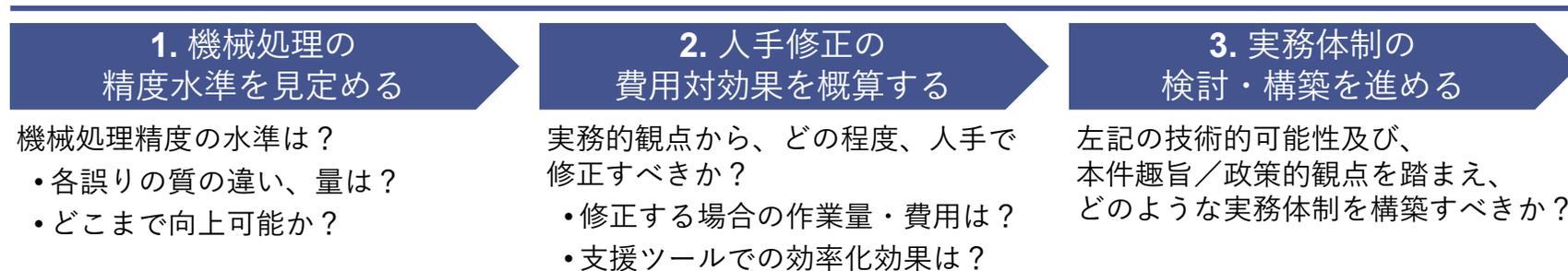


1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
 - ① 実験のアプローチ
 - ② 機械処理の精度水準
 - ③ 人手修正の費用対効果
3. 実務体制構築の方向性オプション（議論用叩き台）

p.3再掲) 実証実験の流れ・位置付け

- 約1,600件の判例を対象に実験を行い、自動仮名処理の技術的な可能性を検証・評価
- 技術的可能性の評価を踏まえ、実務体制構築の検討材料を提示

仮名処理の実務体制を構築するまでの流れ



検討論点



検討／実施アプローチ



本資料にて結果を報告

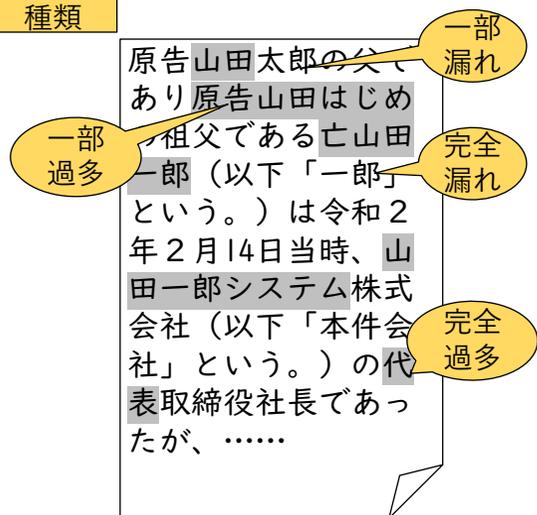
本資料にて検討の枠組・材料を提示

人手修正のオプション

- まず、人手修正の前段の機械処理において、漏れ／しすぎのトレードオフを調整する
- その上で、大枠の修正のオプションは、「判決文を全部見る」か「機械出力だけ見る」かの2つ

機械のトレードオフの調整

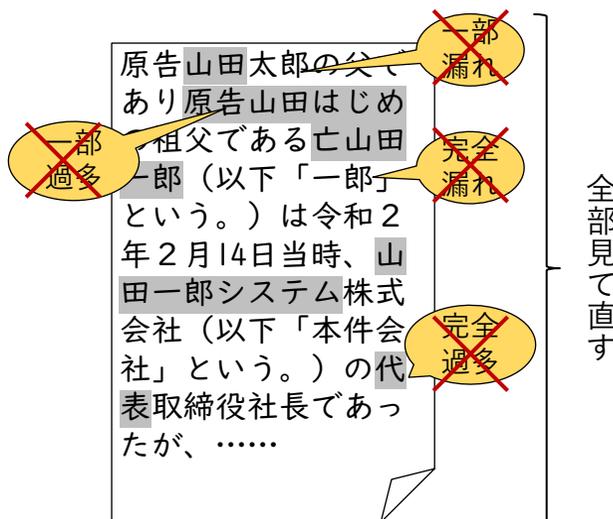
誤りの種類



“漏れ”と“しすぎ”のトレードオフの間でバランスを取る

- “全部黒塗り” ⇔ “何もしない”

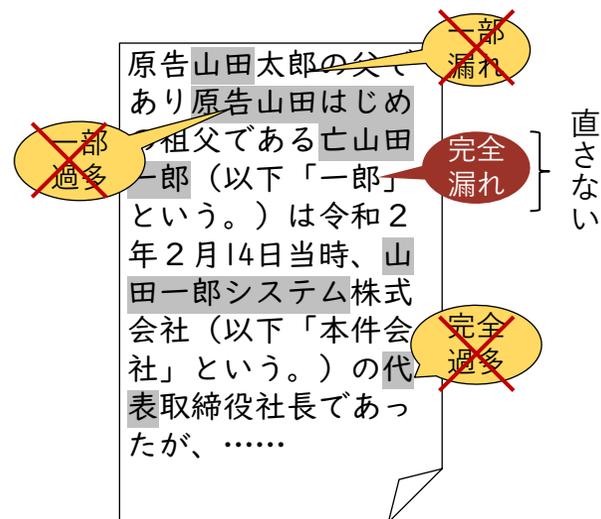
1 判決文を全部見て修正 (全部修正)



完全な仮名を行うために、
判決文全体を確認

- 全部見るが、修正ツールによるアシスト・効率化は可能
(同一語句の同時選択／似た語句のサジェスト、等)

2 機械出力だけ見て修正 (一部修正)



一部の「漏れ」を許容し、機械が
判定した部分の周辺のみを確認

- 機械が判定した部分から、確認対象をさらに絞り込むことも可
(人名／地名／番号、等)
- 機械で「多めに仮名」すれば、「漏れ」割合を一定低減可能

オプションの評価

- 仮名処理に係る費用は大きく分けて、人手処理の人件費と機械処理システムの費用
- 人手処理については、最大で（全文確認修正の場合）、2,500万円前後のイメージ

仮に10万件を想定した場合			1 全文確認修正	2 機械出力のみ修正	3. 人手修正なし	参考
費用イメージ	仮名処理に係る費用	1件あたり費用 ¹⁾ (時間)	250円 (作業時間10分) ※従来作業時間が60分かかる判決を想定	75円 (作業時間3分)	-	機械処理・支援を利用せず、人手で仮名処理した場合は、1件約1,500円程度かかる想定 ※仮名作業時間は約60分を想定
		10万件の場合	2,500万円/年 ※単純計算による概算	750万円/年 ※単純計算による概算	-	
		機械処理システム (開発費+運営費)	(システム費用の考え方/方針検討については、次章記載) <ul style="list-style-type: none"> 機械処理AIの開発・改善・アップデート 作業ツール/システムの開発・保守 			

人手修正作業の
人員体制イメージ

年間10万件の判例を、1日ごとに処理・公開する場合は、**最大でも9人体制**で人手修正を施すイメージ（修正作業量を削減しない場合）

- 1日判例公開数 = 約417件（10万件 ÷ 240日）※営業日ベース
- 9人体制であれば、1日約432件を処理可能（1人が1日で修正可能な判例数 = 48件）

公開を1日ごとに行わない場合（例えば半年に一回）は、都度短期集中的に作業を実施

案件ごとの1/2の組合せや機械処理のチューニング等の個々の改善項目含め、より詳細なオプションは存在するも、次章にて大枠の方向性を定めた上での詳細検討が望ましいと思料

1. 人手作業人員の時給を1,500円と仮置きし、試算

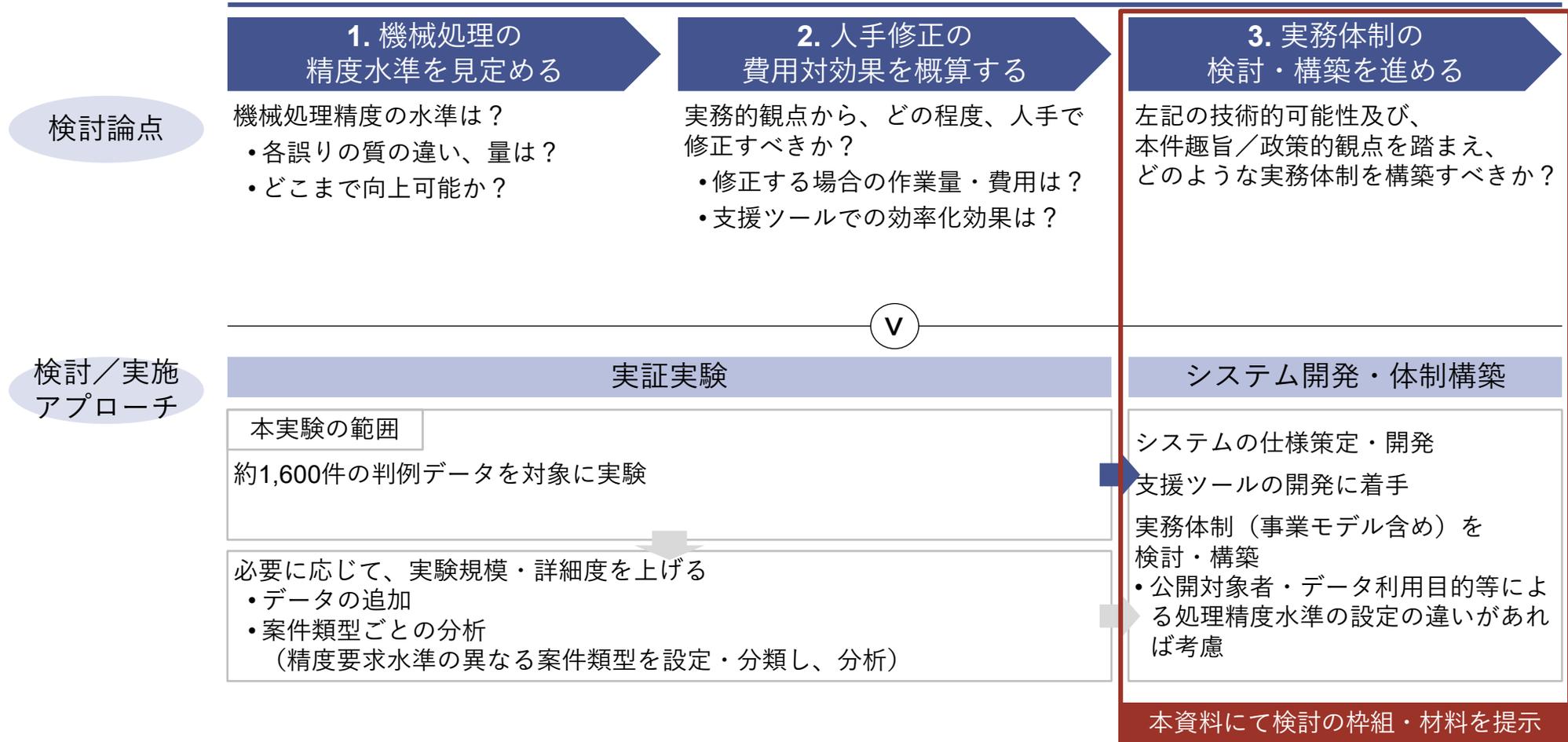


1. 実証実験の概要（目的と位置づけ）
2. 実証実験の結果：自動仮名処理の技術的可能性
3. 実務体制構築の方向性オプション（議論用叩き台）

p.3再掲) 実証実験の流れ・位置付け

- 約1,600件の判例を対象に実験を行い、自動仮名処理の技術的な可能性を検証・評価
- 技術的可能性の評価を踏まえ、実務体制構築の検討材料を提示

仮名処理の実務体制を構築するまでの流れ



技術的可能性及び、本件趣旨／政策的観点を踏まえ、どのような実務体制を構築すべきか？

① 議論の出発点として、「民事判決オープンデータ化」の対象範囲・方式において、**どのようなゴール設定が、選択肢としてあり得るか？**

➤ 全判決公開／一部判決公開、一般公開／特定事業者向け、等

② 各ゴール設定ごとに、必要となるシステム開発費用、準備・調整期間、等の**「取組／検討の“大きさ”」**はどのように異なるか？

➤ 同時に、それぞれにおける残論点／詳細化すべき項目はどう変わるか？

③ 取組／検討の“大きさ”を認識しつつ、本件オープンデータ化の趣旨・政策的位置づけに立ち返った際、**現時点でゴールとして見定め、詳細検討を進めるべき方向性はどれか？**

➤ 今回、重視すべき観点／政策としての評価基準は何か？

➤ 評価基準を踏まえ、選択・詳細化・磨くべき、ゴール／方向性はどれか？

④ 方向性踏まえ、今後詳細検討すべき事項及び検討ステップ・スケジュールは？

本資料では、
①②に関し、
たたき台を
記載

叩き台を元に、
今後の方針を
議論

本資料では、プロジェクト全体のシステム費用／準備・調整期間や、今後の詳細検討項目が変わる分岐点（①+②）を明確化し、議論のたたき台とする

①②実務体制構築の方向性オプション（議論用叩き台）

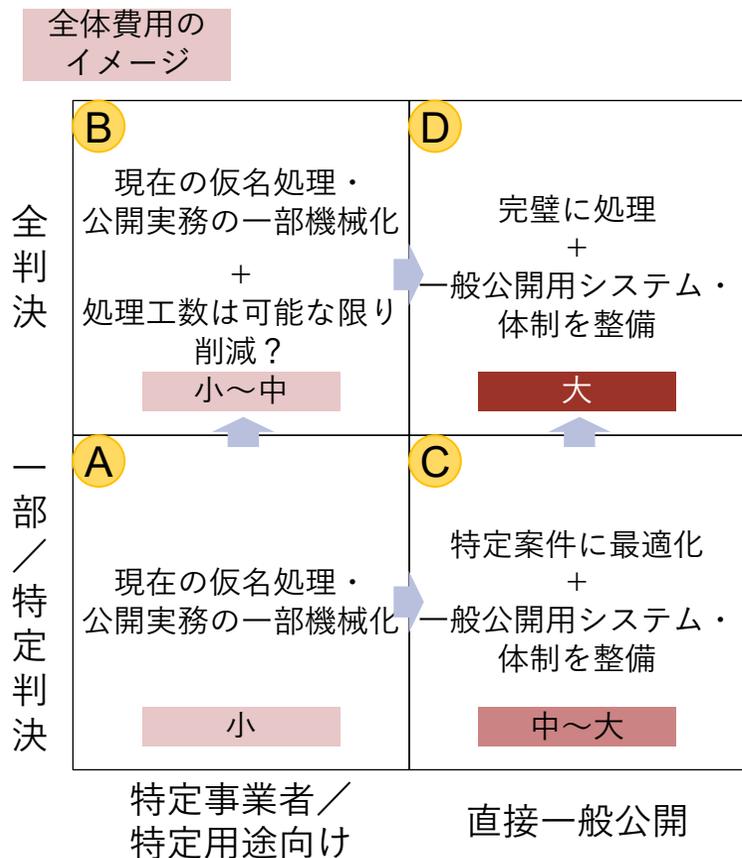
- 従前のPTでの議論を踏まえ、公開対象範囲・方式の組合せとして4つの方向性を仮置き
- 選ぶ方向性次第で、プロジェクト全体の費用／調整期間、検討項目が変わる想定

4つの方向性オプション（議論用叩き台）

各方向性における取組のイメージ

全体費用のイメージ

次に着手すべき検討項目イメージ



A

- 現在、各判例データサービス事業者様にて行われている現行の仮名処理実務を機械処理により効率化／移管

小

- 持続的な座組の検討
- 仮名処理水準は現状と同水準が前提

B

- 基本は現在の処理実務を効率化の方針
- 加えて、全件公開にあたり、処理工数削減余地を重点的に検討

小～中

- 持続的な座組の検討
- 特定案件／公開目的次第での、処理精度低減（修正コスト削減） 余地の確認

C

- 一般ユーザー向けのシステム開発、実務体制（問合せ等）構築の検討が必要
- 特定案件に鑑みた処理工数削減余地を検討

中～大

- 公開目的・範囲の整理
- その上で、必要処理精度／システム要件を詳細化

D

- 一般ユーザー向けのシステム開発、実務体制（問合せ等）構築の検討が必要

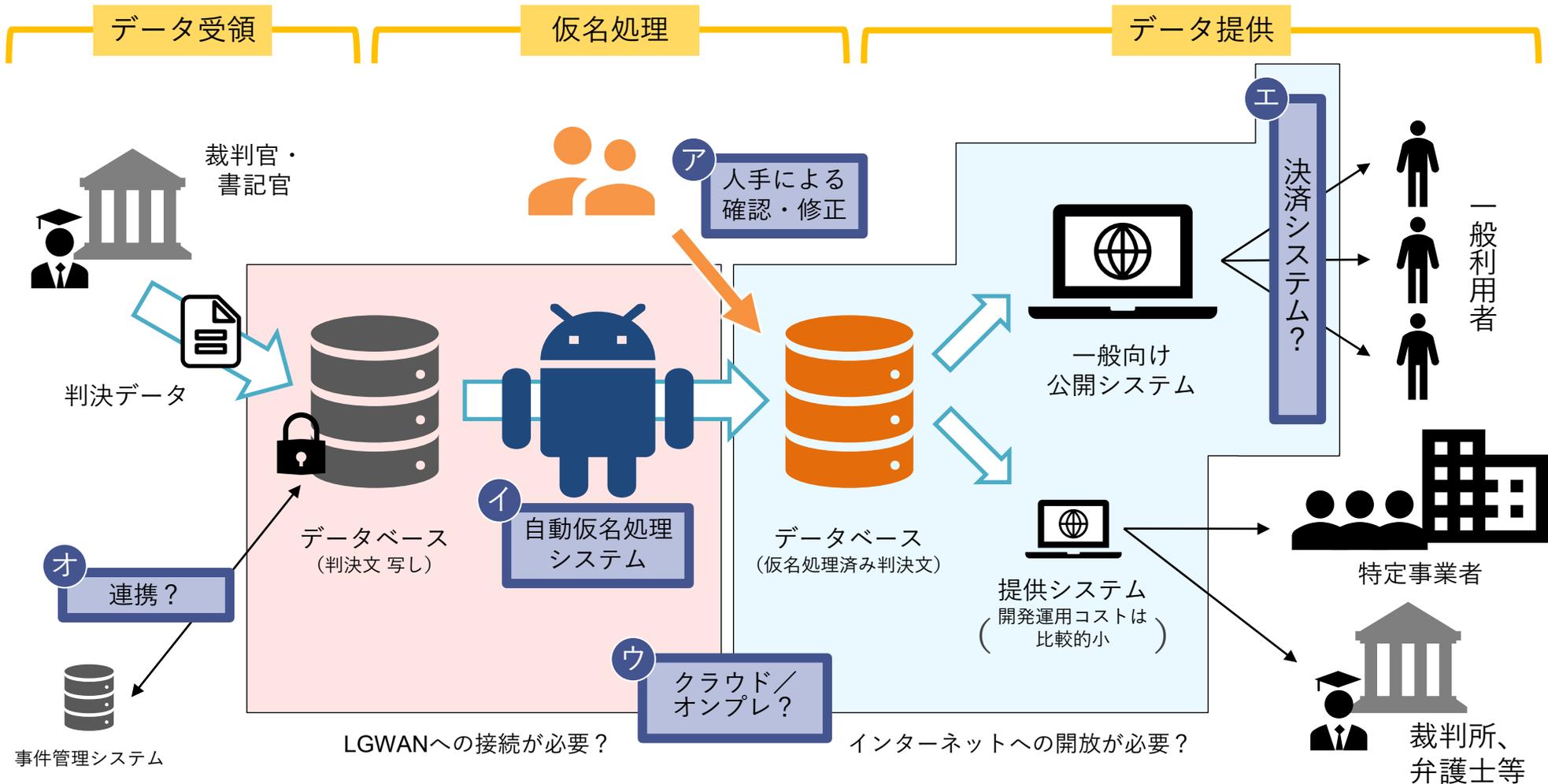
大

- システムの範囲・仕様の検討
- 検討会議体・当事者の整理・調整

※費用の大きめの内訳・大きさ、変動要因については、次頁に記載

②システム費用の規定要因のイメージ (1/2)

ありうるシステム全体像を考えた際、費用を規定する分岐点は、仮名処理関連費用(アイ)の他は、ウクラウドか否か、エ決済システムの有無、オ他システムとの連携要否



②システム費用の規定要因のイメージ (2/2)

- いずれの方向性においても、システムのカバー範囲・仕様次第で、費用は大きく増減
- 公開目的／ゴールの詳細化と、システム要件・費用の具体化を両輪で進める必要

費用を規定する主な要因

A 一部判決→特定 **B** 全件→特定 **C** 一部判決→一般 **D** 全件→一般

費用を規定する主な要因	A 一部判決→特定	B 全件→特定	C 一部判決→一般	D 全件→一般
仮名処理	処理工数は、「処理件数（全件／一部）」×「処理精度」が規定			
	機械処理AI+作業ツール／システム			
データ公開システム	<ul style="list-style-type: none"> • オンプレの場合は、物理的なハードウェアの調達・設置・接続に加え、システム保守に係る人件費を要する • クラウドの場合は、サーバー容量・性能のみに応じて費用が増減 • セキュリティポリシー、等に応じて選択 			
	(データ提供事業者の数や、契約の流動性等次第も、基本は不要なイメージ)		<ul style="list-style-type: none"> • 個人情報・決済情報の管理、決済の実行システムの開発 • 提供対象（利用者）／提供方法、等の座組次第で要否が決まる 	
裁判のIT化全体との連携	(裁判のIT化は、長期的・大規模であり、本件オープンデータ化にて特定事業者・目的向けにデータ公開する場合は、取組対象として劣後するイメージ)		<ul style="list-style-type: none"> • 一般公開にあたり、利用者目線での「裁判システム」全体の整合性・連続性が必要等故、IT化検討との協働・協調が必要な可能性 	

システム関連
金額イメージ

初期
運営・保守

数千万円規模～
(仮名処理費用がメイン)

～数十億円規模
(公開用システムがメイン)

詳細な見積もり（初期／運営・保守費用の切り分け含め）
のためには、より具体的な要件に基づく仕様検討が必要

③④今後の詳細検討事項と進め方イメージ（案）

本年度PTにおいて体制構築の方向性を定め、次年度以降、座組・事業形式／システム要件・仕様含め、詳細化していく進め方が一案

	検討項目	時期イメージ (叩き台)
<p>体制構築の方向性 (現時点) を定める</p>	<ol style="list-style-type: none"> 1. 実務体制構築の方向性オプション (A~D) の内、どれを (優先的に) 詳細検討するかを確定 2. 今後、詳細検討に入るために、どのような会議体・スケジュールを設定するかを検討 <ul style="list-style-type: none"> ▶ 協働すべき関連PT、当事者、等 	<p>本年度中 (本PTにおける結論の位置付け)</p>
<p>方向性・ゴール設定の 詳細化・明確化</p>	<ol style="list-style-type: none"> 1. 公開目的／ゴールの詳細化・明確化と、システム要件・費用の具体化を両輪で進める <ul style="list-style-type: none"> ▶ 費用が大幅に変動し得るシステム要件 (前頁記載) については優先して有無・内容を検討 2. 公開機関の座組・事業形式を検討・確定する 	<p>来年度以降～</p>
<p>実際に、体制構築／ システム開発に着手</p>	<ul style="list-style-type: none"> • 具体化された実務体制方針、システム要件をもとに、予算確定・確保、システム開発を開始 	

参考) 総務省における事例

総務省・法制執務業務支援システムe-LAWSにて、採用されている法令データの作成・編集のためのシステム／インターフェースを参考にできる可能性

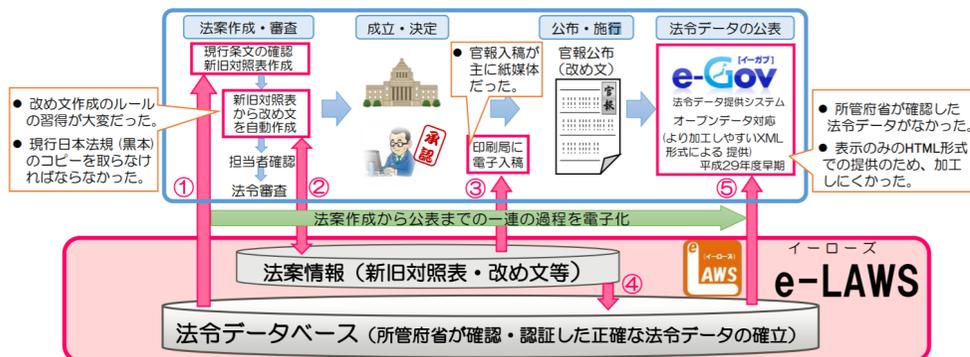
法案作成・編集システム・
インターフェースのイメージ2)

総務省公開資料¹⁾より：e-LAWSの概要

参考 法制執務業務支援システム (e-LAWS) の概要



- ▶ 「霞が関で働く女性有志」の提言を踏まえ、業務省力化・平準化の観点から、ICTを活用し、法案等関係資料の作成支援等を行うシステムの開発を行うことを決定（「国家公務員の女性活躍とワークライフバランス推進のための取組指針」(平成26年10月女性職員活躍・ワークライフバランス推進協議会決定)
- ▶ 総務省行政管理局において、一連の法案等作成業務を支援する「e-LAWS」を開発し、平成28年10月3日から各府省において本格運用を開始
- ▶ ①所管府省が確認・認証した正確な法令データを確立し、法令原本として活用できるデータベースを行政及び国民等へ提供、②新旧対照表から改め文を自動作成するなど、法案担当者の負担を大きく軽減



e-LAWS e-Legislative Activity and Work Support System

総務大臣談話

“e-LAWSの法改正支援ツールで、法令案の立案や参考資料作成の作業を飛躍的に省力化・効率化する”

法令データシステムのデータ形式を 機械可読性の低いHTMLから高いXMLへ変更

e-LAWS導入前 (HTML形式)	導入後 (XML形式)
<pre>行政相談委員会 (昭和四十一年六月三十日法律第九十九号)

 <P> <DIV class="artitle">(目的)</DIV> <DIV class="item">第一条 この法律は、国民の行政に関する苦情の解決の促進に資するため、苦情の相談に関する業務の委嘱について必要な事項を定め、もつて行政の民主的な運営に寄与することを目的とする。 </DIV></pre>	<pre><LawNum>昭和四十一年法律第九十九号</LawNum> <LawTitle>行政相談委員会</LawTitle> <MainProvision> <Article Num="1"> <ArticleCaption>(目的)</ArticleCaption> <ArticleTitle>第一条</ArticleTitle> <Paragraph Num="1"><ParagraphNum/> <ParagraphSentence> <Sentence>この法律は、国民の行政に関する苦情の解決の促進に資するため、苦情の相談に関する業務の委嘱について必要な事項を定め、もつて行政の民主的な運営に寄与することを目的とする。 </Sentence> </ParagraphSentence> </Paragraph> </Article></pre>

ブラウザに表示するための情報 (太字にする、字下げをするなどのみで、法令データの意味情報は無い)

法令データの意味情報を詳細に付与できる
<LawTitle> 法令の題名
<ArticleCaption> 条の見出し
<ArticleTitle> 条名 など

→ 人手でデータを編集・加工

→ アプリケーションでの加工・利用が容易に

※画像は日経コンピュータ記事²⁾より

法令作成者は、Wordのような直感的な編集ソフト／インターフェースによりXMLを編集

- 作成者は、XMLタグが直接見えない形式で法令データを作成・編集可能
- 編集結果をもとに（改正前後の法令の差分から）自動的に一部改正法令（改め文）を生成

→本件においても、Word等で書く際に、ハイライトやフォント変更のような動作で仮名対象を指定し、仮名化部分がタグ付されたデータを生成できる可能性



1. 総務省報道資料「法制執務業務支援システム (e-LAWS) の本格運用開始に際しての総務大臣談話」 https://www.soumu.go.jp/main_content/000441588.pdf
2. 日経コンピュータ 2017年8月31日号 pp.44-45「紙の聖域に切り込む 戦後初、行政・司法のデジタル改革」

End of Document