

Confidential

民事判決データベース化事業の在り方に関するWG

実証実験の結果報告

－仮名処理の実務体制構築に向けて－

2022年3月30日
株式会社Legalscape

背景

民事判決のオープンデータ化にあたっては、判決の仮名処理実務の持続的かつ経済的な体制構築が必要である

昨年（'20年度）の実証実験では、機械による仮名処理の技術的可能性を評価した

- 機械処理精度は一定の水準を達成し、実務体制への採用・組み込みは充分検討に値することが確認された
- 同時に、機械処理後の人手によるチェック・修正の要否、必要な場合の人員・実務体制の在り方について、更なる検討を要することも認識された

⑤

目的

本年度の実証実験は、人手作業を含めた実務体制を構築する場合の検討論点を整理した上で、想定される業務フローの策定・人手作業ツールのUI設計、等を行い、来年度以降、実際に体制構築に着手するにあたっての方針検討に資することを目的とする

第1回WG (7/14)

資料5 (p.2) より

事業化WG・検討事項(含む予算)の概要(イメージ)

① 民事判決情報の取得体制の構築

- ・取得対象とすべき民事判決情報の範囲(調書判決を含めるか等)、付加情報の取得の要否、可否
- ・取得する民事判決情報のデータ形式(PDF、ワード、テキストデータ等)
- ・取得のフローの構築(裁判所の事件管理システムとの連携の可否を含む。)

② 情報管理機関の在り方

- ・情報管理機関となる組織の適格性
- ・適格性維持のため必要となる人的・物的体制の在り方

③ 仮名処理の実務体制の構築 ※実証実験の要否、内容

- ・自動仮名処理システム、人手修正支援ツール、仮名化前後のデータベースの開発
- ・機械処理のフローの構築
- ・人手修正のフローの構築、既存の民間事業者との連携

④ 民事判決情報の保管管理体制の構築

- ・仮名化前後の民事判決情報の保管管理方法(クラウド/オンプレミス、保存期間等)
- ・情報セキュリティ体制の構築等

⑤ 民事判決情報の提供システムの構築

- ・提供システムに実装すべき機能
例えば、検索機能(言渡裁判所、年月日、事件番号、フリーキーワード機能、事件類型別等)
- ・提供のフローの構築、料金体系(件数ごと、容量ごと、サブスクリプション制等)、費用感

⑥ 民事判決情報の提供のための要件等

- ・利活用機関の範囲、要件
- ・利活用機関に対する提供の条件の要否、内容(情報セキュリティ体制、目的外利用の禁止等)
- ・ビッグデータとしての利活用を目的とする場合の提供の在り方

⑦ 言渡から提供までの期間

- ・利活用機関のニーズを踏まえた期間設定、早期提供のためのフローの構築

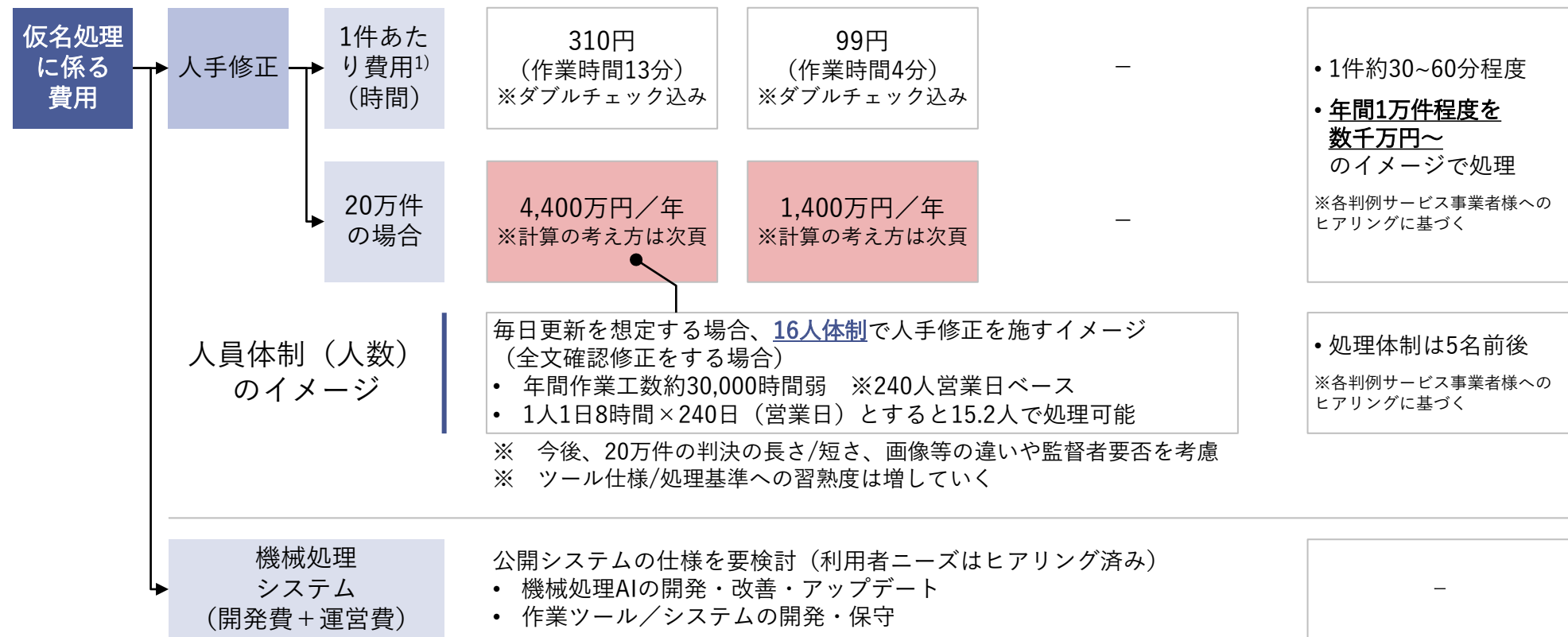


1. 実験内容の報告：人手修正作業ツールの開発と検証
2. 関係者ヒアリング内容の共有
3. 今後の検討方針（再掲・一部改変）

実証実験の結果（概要）

- 仮名処理に係る費用は大きく分けて、人手修正の人件費と機械処理システムの費用
- 仮に20万件を全文確認修正する人件費は、年4,400万円程度が一つの目安

仮に20万件を想定した場合



評価の観点：体制構築の現実性（コスト） + 現状の各社仮名実務の集約による経済性

GALSCOPE

WE EMPOWER LEGAL PROFESSIONALS

1. 人手作業人員の時給を1,500円と仮置きし、試算

コスト推定の考え方

本実験に使用した判決と、実際にオープンデータ化の対象となる判決の質・量を考慮して、なるべく保守的な推定を実施

実験対象データ

- ✓ 約1,600件の判決データ
 - ✓ 東京地裁
 - ✓ 平均文字数8,500文字
 - ✓ 図表・別紙なし
- ✓ このうちの約100件を使用
 - ✓ 学習とテストへの振り分けはランダム実施
 - ✓ 各判決に対し1-3回程度、一連の実作業~チェック作業を実施し、時間等を計測
 - ✓ 平均文字数4,500文字

判決文の文字数を考慮し、1件あたりの作業時間・コストを推定

オープンデータ化の対象となる判決¹⁾ = 約20万件

第一審判決数	全体	(内対席)	(内欠席)
民事通常訴訟事件 (地裁)	57,549	32,730	24,779
民事通常訴訟事件 (簡裁)	139,843	42,996	96,831
行政訴訟事件 (高裁)	207	190	17
行政訴訟事件 (地裁)	1,468	1,313	152
合計	199,067	77,229	121,779

控訴事件	既済件数
民事訴訟事件 (高裁)	12,228
行政訴訟事件 (高裁)	777
民事訴訟事件 (地裁)	3,944
合計	16,949

上告・上告受理申立事件	既済件数
民訴訟事件 (最高裁)	3,762
行政事件 (最高裁)	763
民事訴訟事件 (高裁)	445
合計	4,970

20万件の内訳

※欠席判決は、訴状を引用するだけの判決文になるため、対席判決より短い

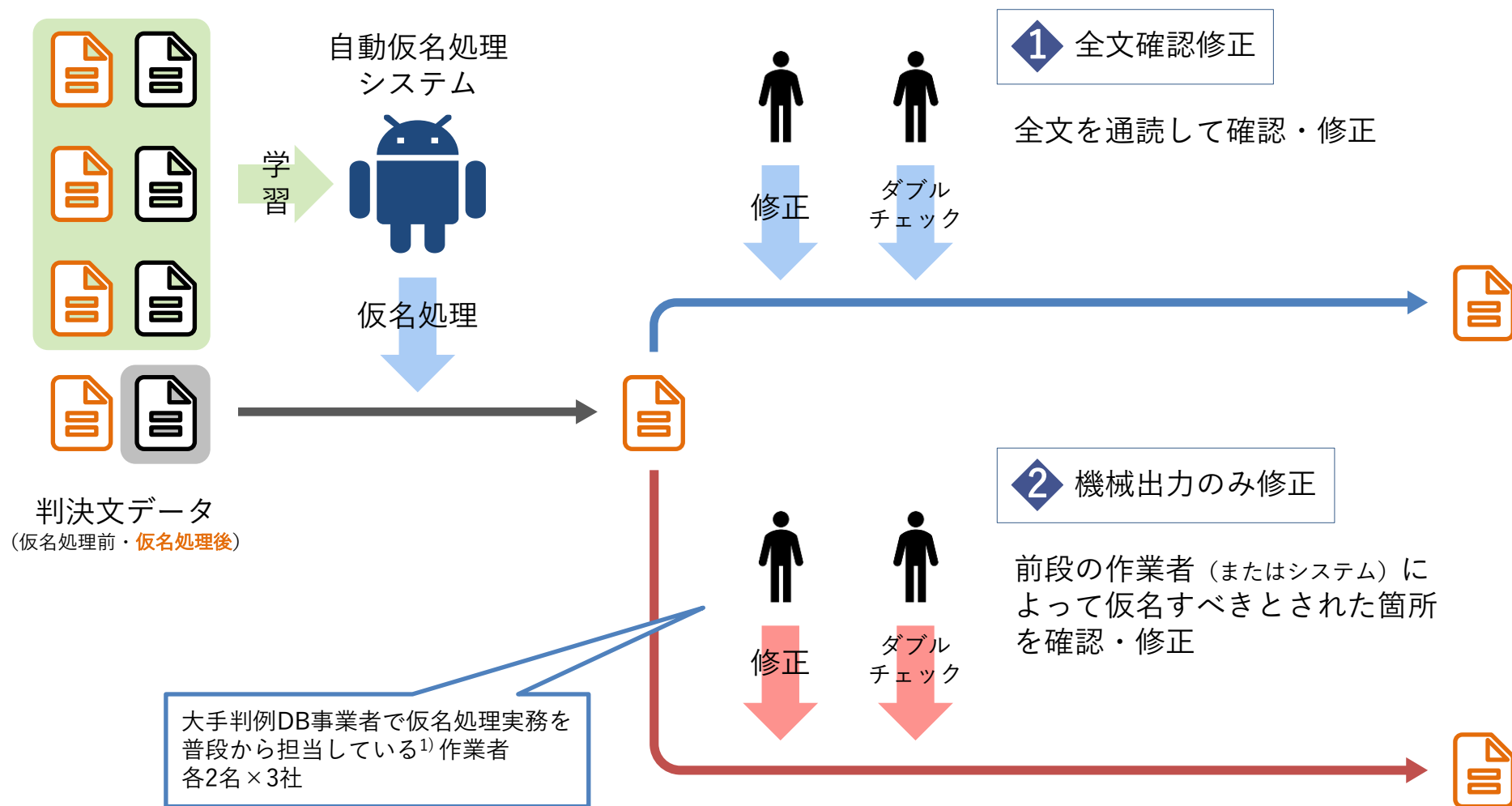
参考) 「既済」件数 (和解+判決)

公開対象の約20万件のうち、欠席判決12万件は、実験結果の半分の時間 (保守的)、と仮定し、全体の作業コストを推定

1. 裁判所HP公開情報より集計 https://www.courts.go.jp/vc-files/courts/datebook/2021/db2021_35p-54p.pdf

実証実験の内容

6名の参加者でのべ150件の仮名処理を行い、2シナリオで時間計測（詳細は別紙参照）



1. 条件は異なる。今回の実験では仮名基準として第一法規株式会社の基準で統一し、また弊社開発の仮名処理作業専用ツールを使用

実証実験の結果（概要）

- ツールの体感的な利便性は高い評価
- 処理時間についても、現状比で大きく効率化された

結果

	1 全文確認 修正	2 機械出力 のみ修正	現状（ヒアリングに 基づく参考値）
作業	7.2分	2.8分	
チェック作業	5.2分	1.1分	
合計	12.5分	3.9分	（平均的には） 30-60分程度

文字数考慮
後の時間

主な 傾向

（詳細は
次頁以降）

【全般：ツールへの評価】

- 使用者所感としては、体感数倍以上、非常に効率化できている
 - AIの精度の高さに驚いた
 - 仮名作業自体のツールの使いやすさも、直感的で効率的。マニュアルを参照する場面がなかった

【全文確認修正の場合】

- 判決文が長いほど/仮名対象が多いほど時間がかかる
 - 機械処理の「精度」と作業時間の相関は不明確
- 今後の処理時間短縮は、機械処理精度というよりはツール効率+熟練度UPがメインか

【機械出力のみ修正の場合】

- 仮名対象数と「精度」のみが作業時間に寄与（全文確認では判決が長いほど作業時間を要したのに対して）
- ダブルチェック作業は判決によらず安定して早い

実証実験の結果（傾向）

事件類型¹⁾によって判決の長さや仮名対象数が異なり、結果時間も変動

	割合	文字数 (平均)	仮名対象数 (平均)	F1値 ²⁾ (平均)	合計作業時間 ³⁾ (平均)	
					1 全文確認 修正	2 機械出力 のみ修正
損害賠償請求事件	23.8%	8,288	8.9	0.905	10.8分	1.7分
貸金請求事件	15.9%	3,787	7.8	0.901	6.2分	1.6分
発信者情報開示 請求事件	14.3%	6,737	10.8	0.878	12.4分	3.1分
その他	46.0%	3,389	7.8	0.963	5.6分	1.4分

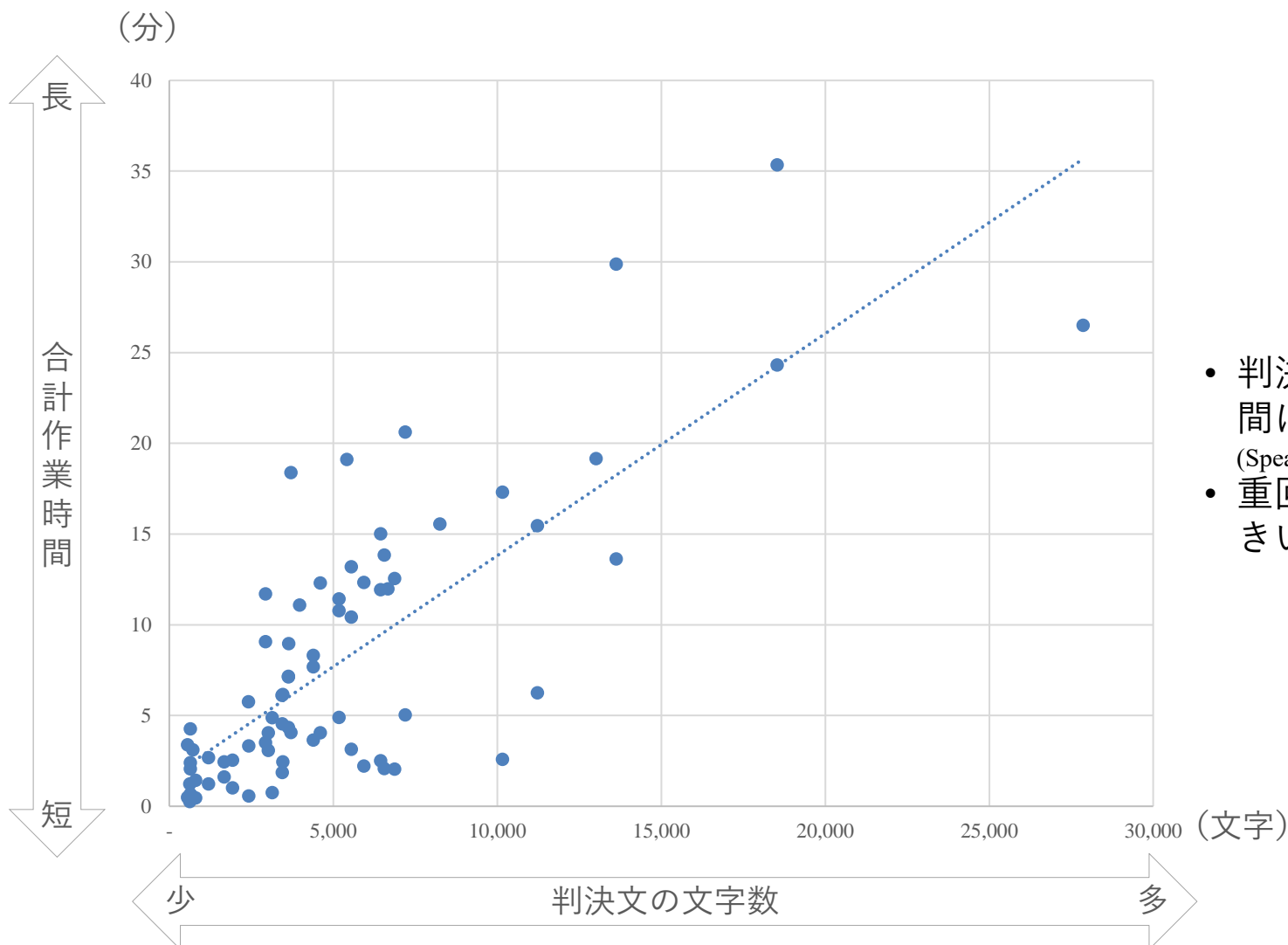
1. D1-Law.com 上での事件名をもとに、「等」がつくだけなど十分近いと考えられるものはまとめてある

2. 説明は後述

3. 合計作業時間は修正作業とダブルチェック作業の時間の合計

実証実験の結果（傾向）

1 全文確認修正 では判決文が長いほど作業時間が長くなる傾向

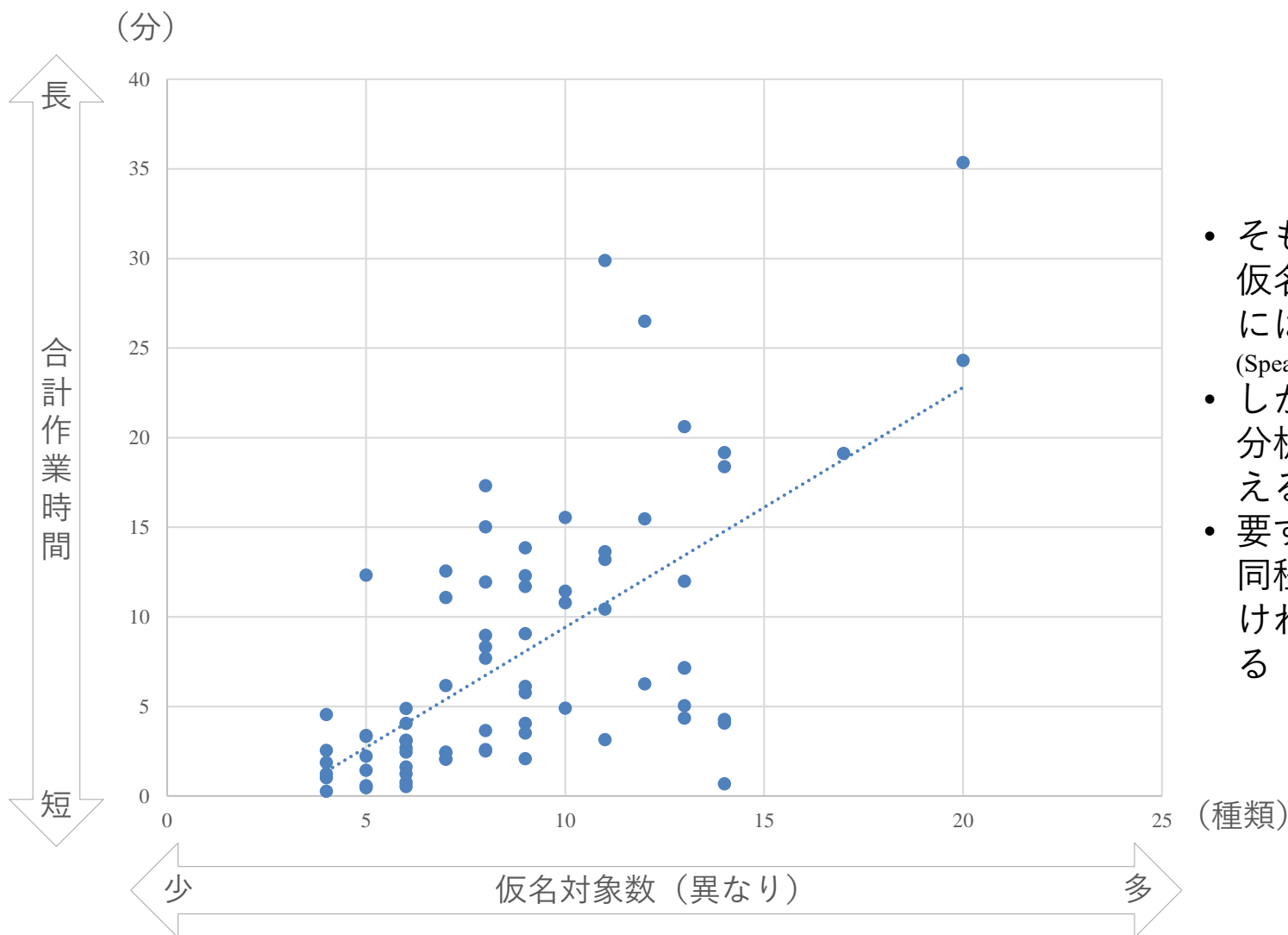


- 判決文の文字数と合計作業時間には強い相関が見られた (Spearman's $\rho = .722$)¹⁾
- 重回帰分析²⁾ から寄与は大きいと言える ($t = 7.65, p = 1.1 \times 10^{-10}$)

1. Spearman's ρ は相関係数の一種で、 $-1 \leq \rho \leq 1$ の値をとる。絶対値が大きいほど強い相関を示す
2. 文字数、仮名対象数（異なり）、F1値、およびダミー変数（被験者、事件類型）を目的変数とした ($p = 3.8 \times 10^{-23}$)。合計でなく各時間においても傾向は同様だった

1 全文確認修正

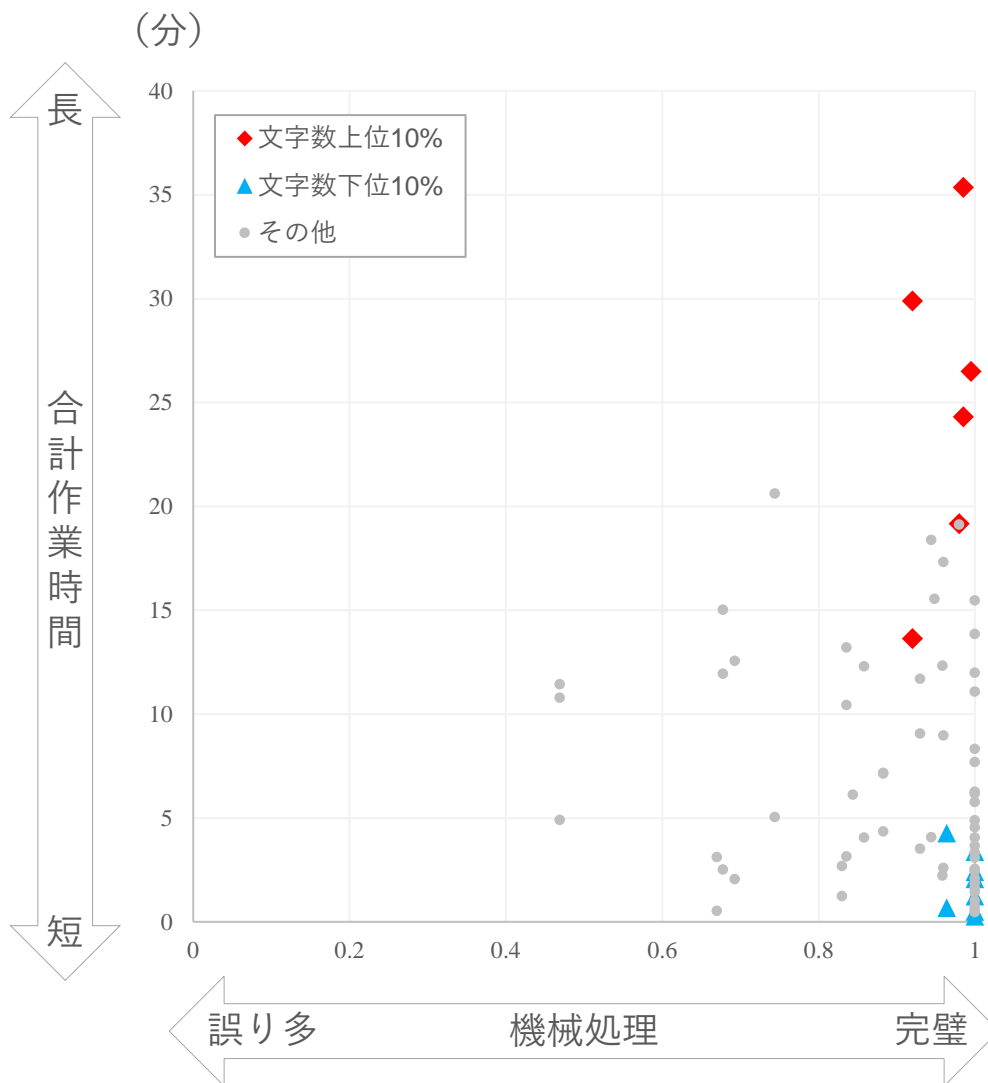
では仮名対象が多いほど作業時間が長くなる傾向



- そもそも判決文の文字数と仮名すべき対象の種類の数には強い相関が見られた (Spearman's $\rho = .697$)
- しかしいずれにせよ重回帰分析から寄与は大きいと言える ($t = 4.43, p = 3.6 \times 10^{-5}$)
- 要するに、判決文の長さが同程度でも仮名対象数が多ければ時間はかかると言える

実証実験の結果（傾向）

1 全文確認修正 では機械処理の「精度」と作業時間の相関は不明確 (F1値)



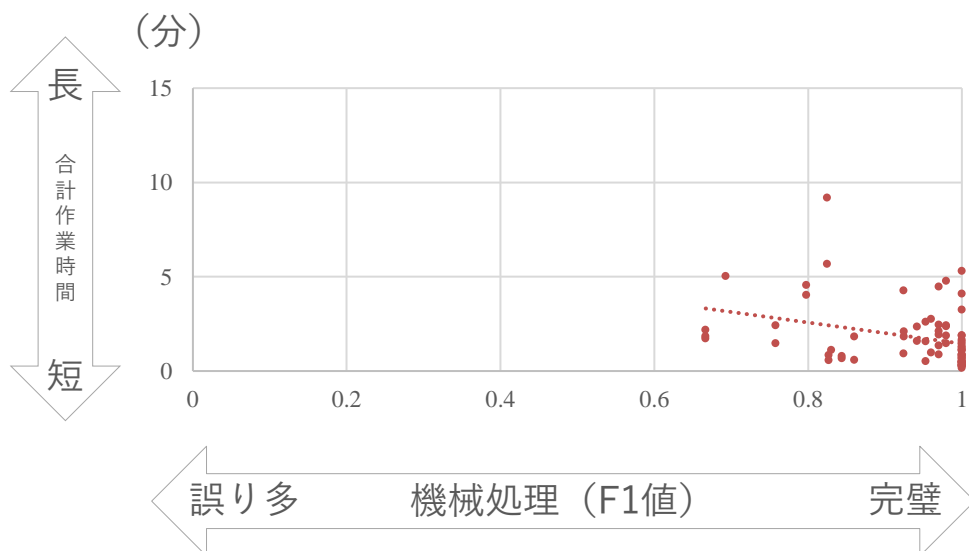
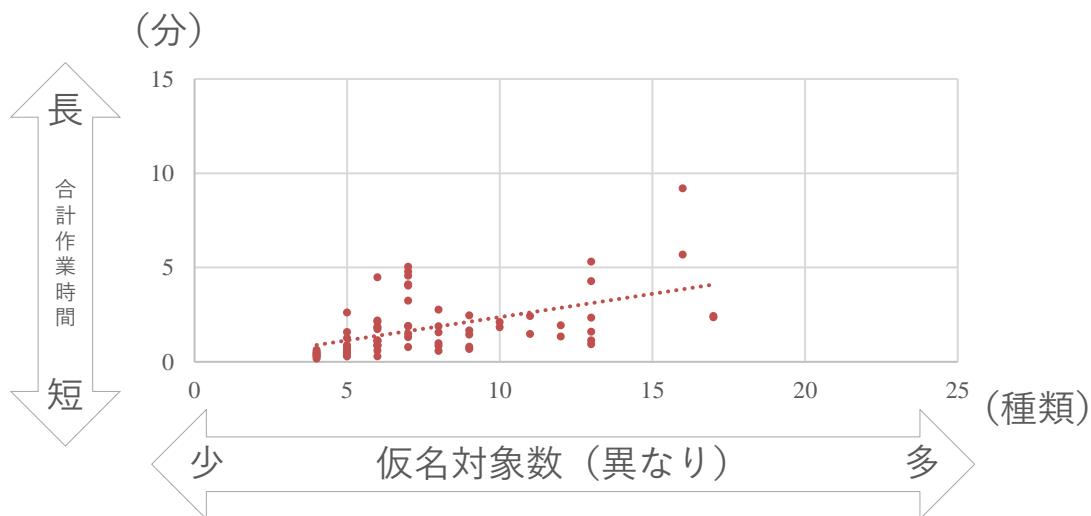
重回帰分析としては「無関係」 ($t = -0.38, p = 0.7$) で、
実際F1値が同程度でも作業時間は短いものから
長いものまで幅広く分布しているが

- そもそも機械処理が完璧に近い判決文が大半であり、誤りが多いケースのデータがほぼない
- 唯一の正解を前提とした採点であるが、そもそも人間の判断も割れたり、複数の正解が存在したりすることも考えられる

横軸：F1値

機械処理による仮名処理を、第一法規株式会社提供の人手による仮名処理を絶対的な正解として採点した点数。0から1の値をとり、正解と完全一致で1となる仮名過多・漏れをバランスよく減点する、適合率と再現率のある種の平均

② 機械出力のみ修正 の修正作業では仮名対象数と「精度」のみが寄与

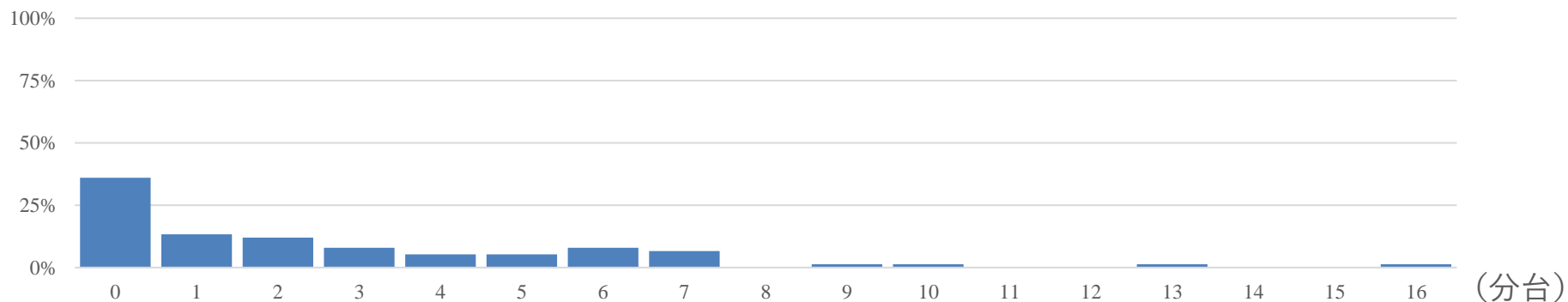


- 重回帰分析から、仮名対象数が多いと時間がかかり、機械処理のF1値が低いと時間がかかると言える ($p = 1.4 \times 10^{-4}$, 3.4×10^{-2} , resp.)
- 一方判決文の文字数単独では時間には影響がなく、長大な判決文に対しても一定の時間内で作業を終えることができると言える ($t = -0.13$, $p = 0.9$)
- ※ ただし実際には文字数は仮名対象数と相関あり

実証実験の結果（傾向）

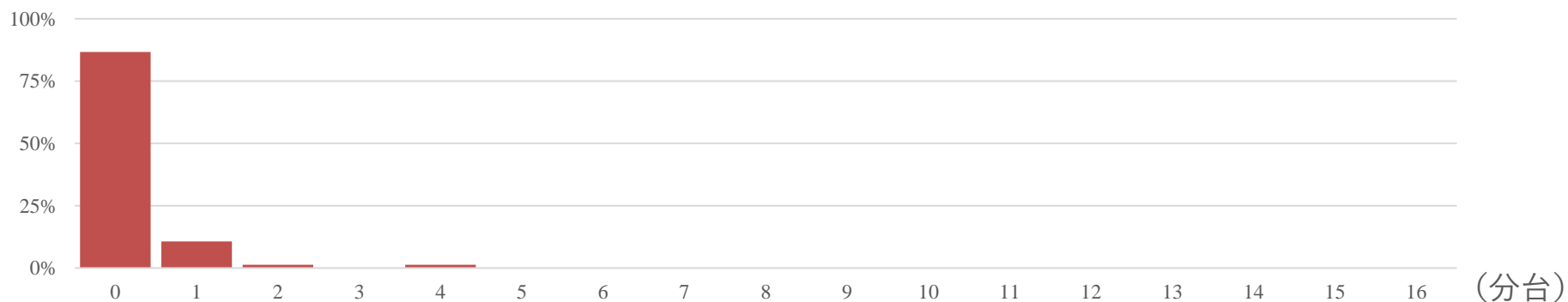
② 機械出力のみ修正 のダブルチェック作業は判決によらず安定して早い

① 全文確認修正



② 機械出力のみ修正

※ 文字数、仮名対象数、F1値、いずれも重回帰分析で $p > 0.05$



短

合計作業時間

長



1. 実験内容の報告：人手修正作業ツールの開発と検証
2. 参考) 関係者ヒアリング内容の共有
3. 今後の検討方針（再掲・一部改変）

1

実務体制構築の方針 に関するご意見

- 主に現在の仮名処理を
されている方・事業者様から

2

オープンデータ化の形式 に関するご意見

- 主に利活用機関・利用者
の方・事業者様から

オープンデータ化事業の方向性について、
実務の実現性と、ニーズの両面からヒアリングを実施

1 実務体制構築の方針に関するご意見

実験に参加された各社様の実感としては、機械処理を組み込んだ実務フローの効率性・実現性は一定高いが（問い合わせフローも含め）、やはり仮名化基準の設定が肝になりそうだ

機械処理の可能性 について

- 実験前は機械処理がどの程度のレベルか不安だったが、実験を通じて「ここまで高いのか」という驚きと一定の安心感を持った
- 作業ツールの使いやすさ・効率も踏まえると、現状の実務を数倍以上時間短縮できている
- ただし、20万件という桁の違う量を処理することを想定した場合には、少しでも時間短縮をしたいということだろう（ツールの文字サイズや行間等含めた細かい改善も必要そう）

実務体制構築に向けたご意見・知見

参考）仮名化基準の明確化が必要だ（仮名化WGと協働する必要）

- 各社の現在の基準に照らして乖離しすぎないように、なるべく網羅するように作ってほしい
- 情報管理機関側で基準が定まったら、基本的にはその基準で利活用機関側でも公開したい
- 推知情報をどこまで仮名するかで人手の作業量が増減する（大学の学部名、コンビニの支店名、病院の部署名）
- ニーズとしてはあるのが、下級審との仮名記号の平仄をそろえること

仮名結果への利用者からの問い合わせ

- 現状は年数件～10件程度、半分が本文自体に関するもの（誤字脱字、更正決定の有無）
- 2割程度が当事者（弁護士）からの外してくれという依頼
 - 無罪判決が出たときにその人の名誉のために仮名を外してほしい、等
 - 当事者であることを確認したあとに対応
- 3割程度が仮名付与の希望（民事事件の企業名、敗訴側の弁護士名、等）
- 裁判所からの更正決定の反映も必要

② オープンデータ化の形式に関するご意見

オープンデータ化の形式について、足元の利用者・利用用途を想定した場合、ニーズ面から見ても、ニーズに沿った公開形式（検索体験・公開スピード）の実現は一定可能そう

想定される利活用 機関と用途

現状と同様、判例データサービス事業者や判例雑誌社等が中心になりそう

- 編集企画という観点からは、学問的な検討をする上では原本が見られたほうがよいだろう
- 労働事件での当事者の性別、配転（転勤）先の住所地、呼び名の加害性（「ノロマ」など）など

新興のリーガルテック企業などが「先例として」活用しはじめるのは、オープンデータの対象が過去分はなく、これから公開されるものだけだとすると、公開後1、2年というよりはある程度蓄積されたあとになるだろう（弁護士が先例として使うには過去10、20年が必要だろう）

- 先例として使う以外には、全量ある場合は統計的な活用ができる可能性はあるが、ニーズ・事業性などは未知数

希望する公開 形式

検索・ 絞り込み

基本的には、年／事件名／関連条文で絞り込みたい

その他、企業法務部分野の判決（原告か被告が企業である判決）に絞って取得したい場合がある

- 法律での絞り込みは一定有益だが、基となっている法律での検索・絞り込みは、例えば709条（損害賠償）だと、企業には限らないのもう少し絞れたら嬉しいが、そこは利活用機関で仕分けることかもしれない
- 絞り込む場合は全件いくら、ではなく絞った場合の価格体系があると嬉しい

事件名については当事者（原告）が決めているが、裁判所側で決めてもらうのが良いかもしれない

公開 スピード

利用者目線だと、現状紙ベースでは2、3か月、DBでは1、2か月で公開されているイメージ

利活用機関として執筆中・企画中の案件がある場合は、関連判決が出たらすぐほしい、1週間程度で入手できたら非常にうれしい

1. 実験内容の報告：人手修正作業ツールの開発と検証
2. 関係者ヒアリング内容の共有今後の検討方針（再掲・一部改変）
3. 今後の検討方針（再掲・一部改変）

今後の詳細検討事項と進め方（大枠）

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（次頁記載）、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>昨年度実施済み</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>本年度は、 1.の人件費の推定・作業ツールの設計を実施</p> <p>来年度以降、 残検討事項に着手</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 	<p>来年度以降～</p>

振り返り) 実務体制構築の方向性オプション (たたき台)

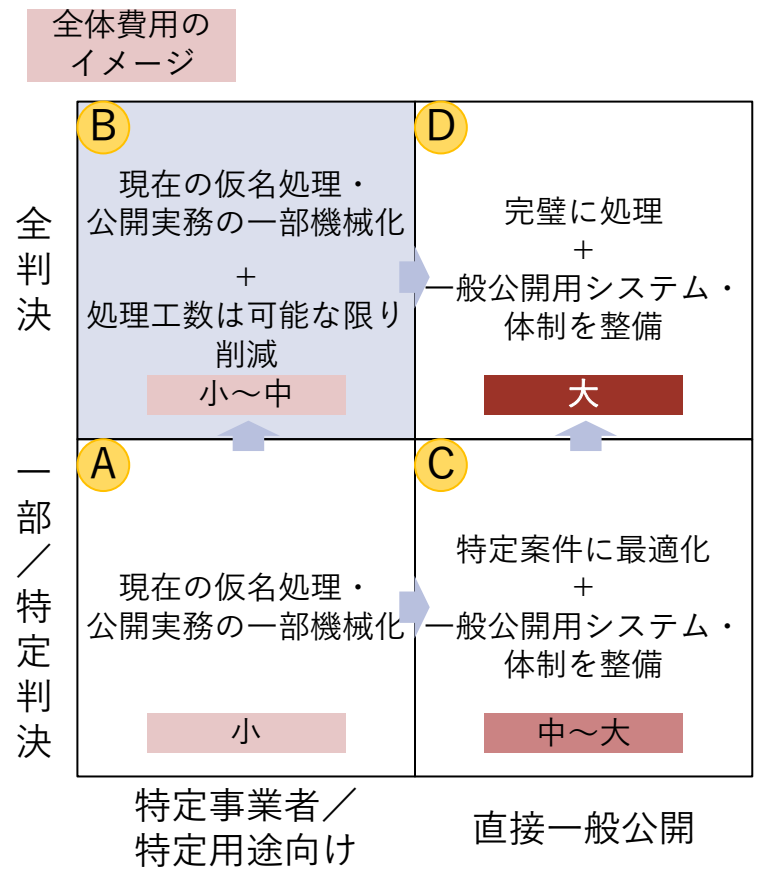
- 従前のPTでの議論を踏まえ、公開対象範囲・方式の組合せとして4つの方向性を仮置き
- 選ぶ方向性次第で、プロジェクト全体の費用/調整期間、検討項目が変わる想定

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
- 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
- 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
- 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

全体費用のイメージ

小

小～中

中～大

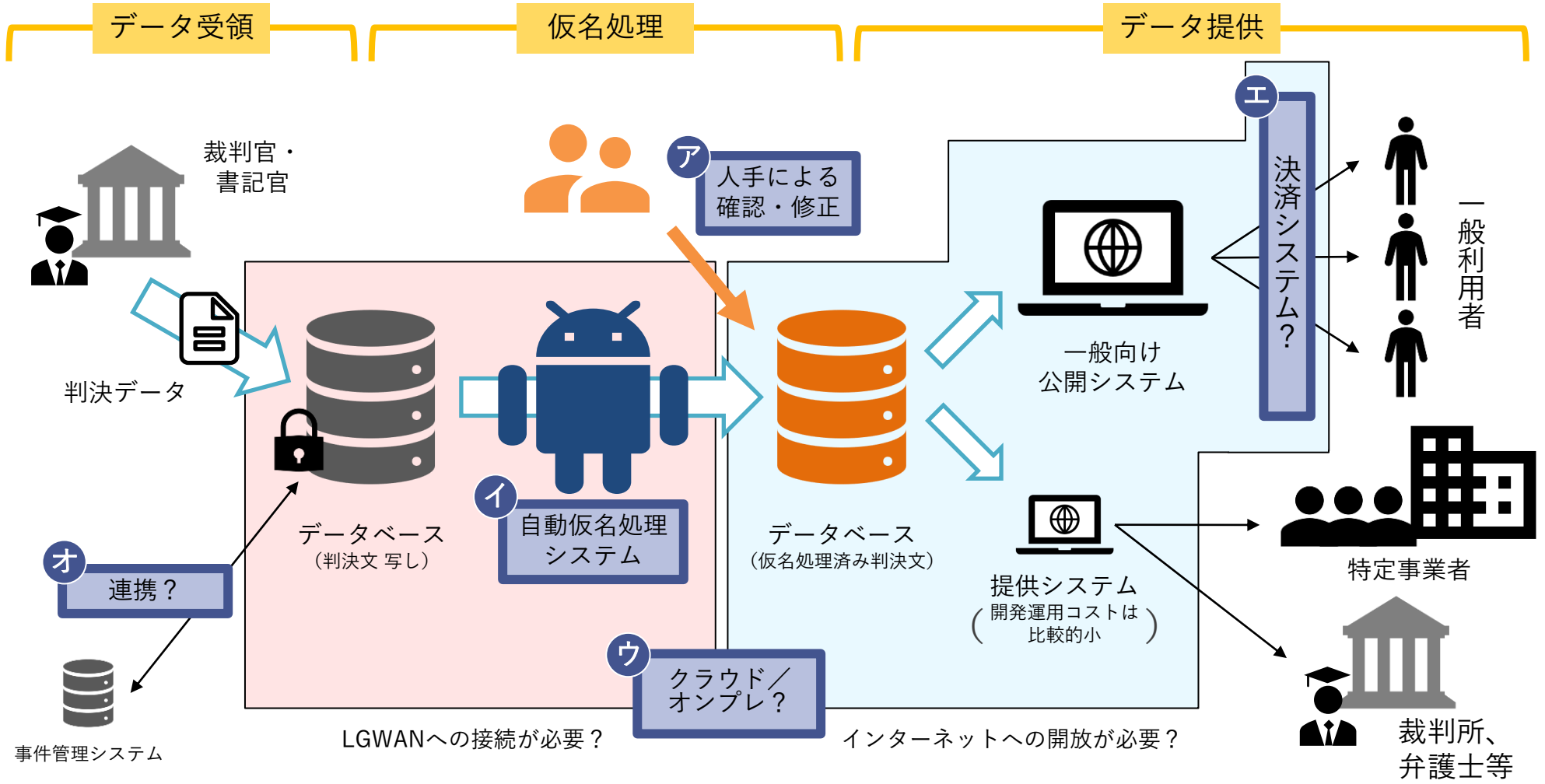
大

- 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
- 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
- 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
- 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

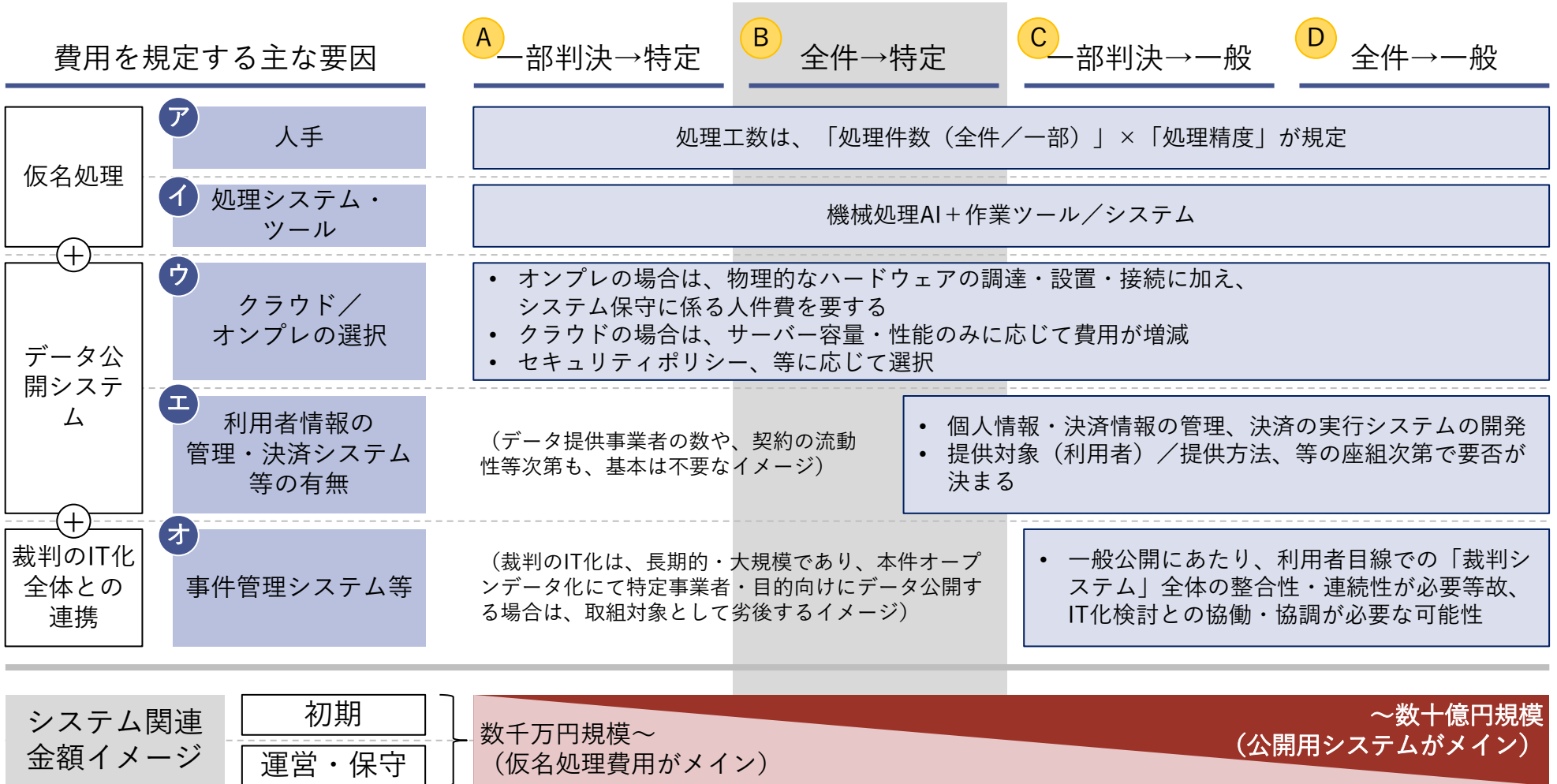
振り返り) システム費用の規定要因のイメージ (1/2)

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



振り返り) 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



End of Document