

## AIを活用した知の共有化の進展への期待 【要約】 【2017年5月23日】

### はじめに

図書館サービスの将来方向として、出版物に限らず様々な情報機関が保有する文化情報資源を一元的に活用できるようにして、新たな知識を創造する「知識インフラ」としての方向性を検討してきた

その方向性の実現に、メタデータの集約による単なる統合検索が可能な「知識インフラ」でなく、情報資源そのもの（画像データ、本文フルテキスト等）をビッグデータとして活用して、人工知能により知識化し、ファクトを取り出せる基盤としての「真の知識インフラ」の構築の実現を見通せる時代になった

「電子図書館」は、「真の知識インフラ」として、「情報の集合体としてのアーカイブ」に留まらず、「人工知能により知識化」され、利用者に対して「電子図書館」を「脳の外部記憶」として利用を可能にする実用化システムの構築も実現可能なところまで到達しつつある。

そのような時代に、図書館を含めた文化情報資源の保有機関は、どんな業務・サービスに人工知能を活用すべきか、そのためにどんな知識とスキルを持つべきかを考察する

### すでにデジタルトランスフォーメーション(デジタル変革)は始まっている

【参照】IT人材白書2017【2017年4月IPA】

今は、IoT、ビッグデータ、ロボット、AI等の技術革新による、第4次産業革命の入り口にいる

あらゆるものがインターネットに接続するIoTの広がり、あらゆる情報がビッグデータとして活用され、AI技術により、様々な分野で定型的な業務はもとより、人海戦術では不可能だった業務まで、AI技術を適用したサービス、ロボットの適用が始まっている

既存のビジネスや業務に新技術を取り入れるだけでなく、ビジネスモデルを変え、経済活用のみならず、個人の生活や社会構造まで影響が及ぶ

デジタルフォーメーション（デジタル変革）とは、あらゆる情報がデジタル化され、IT技術によって、社会や産業、企業、人のあり方や働き方が変わっていくこと

第4次産業革命が進むにつれて、発展するビジネスと縮小するビジネスが明確になっていく

時代環境が大きく変わる時、それにそぐわないビジネスは淘汰されていく

匠の技術的な高度な伝統的技能を要する作業や、旧来の延長線で仕組みの高度化、洗練により、生き残れるビジネスもあるが、現状維持のビジネスの多くは、相対的に意義を失う可能性が高い

IoT、ビッグデータ、ロボット、AI等の技術を、クラウドコンピューティングやモバイル環境で活用できるようになったことは、少ない投資で事業を立ち上げることが可能であり、中小企業、ベンチャー企業や個人の活躍のまたないチャンスである

#### 組織として、

時代の潮流を捉えて、組織が社会の変化の中で、時代に適合して発展できる道を探り、ビジョンをはっきり示すことが重要であり、それは経営者の責務

「デジタルトランスフォーメーション」を実現するには、ビジネスとデジタルのスキルを併せ持った人材の育成と獲得をしていく必要がある

#### 個人として、

自らも「デジタルトランスフォーメーション」の流れの中にあることの意識

求められるのは、周囲を巻き込みながら改革を進める能力やビジネスとデジタルを結び付けて全体をデザインする能力を持った人材になること

目の前の業務だけにとらわれることなく、広く視野を持って進むべき道を探り、学ぶ。勉強会やコミュニティなど、学びの場は周囲にある。自己研さんによって能力を高めれば高めただけ、社会をリードしていく人材になっていく

### 本格的な人工知能の実用化を迎えて

電子図書館は、図書館界、出版界におけるデジタルフォーメーションである

#### 今まで

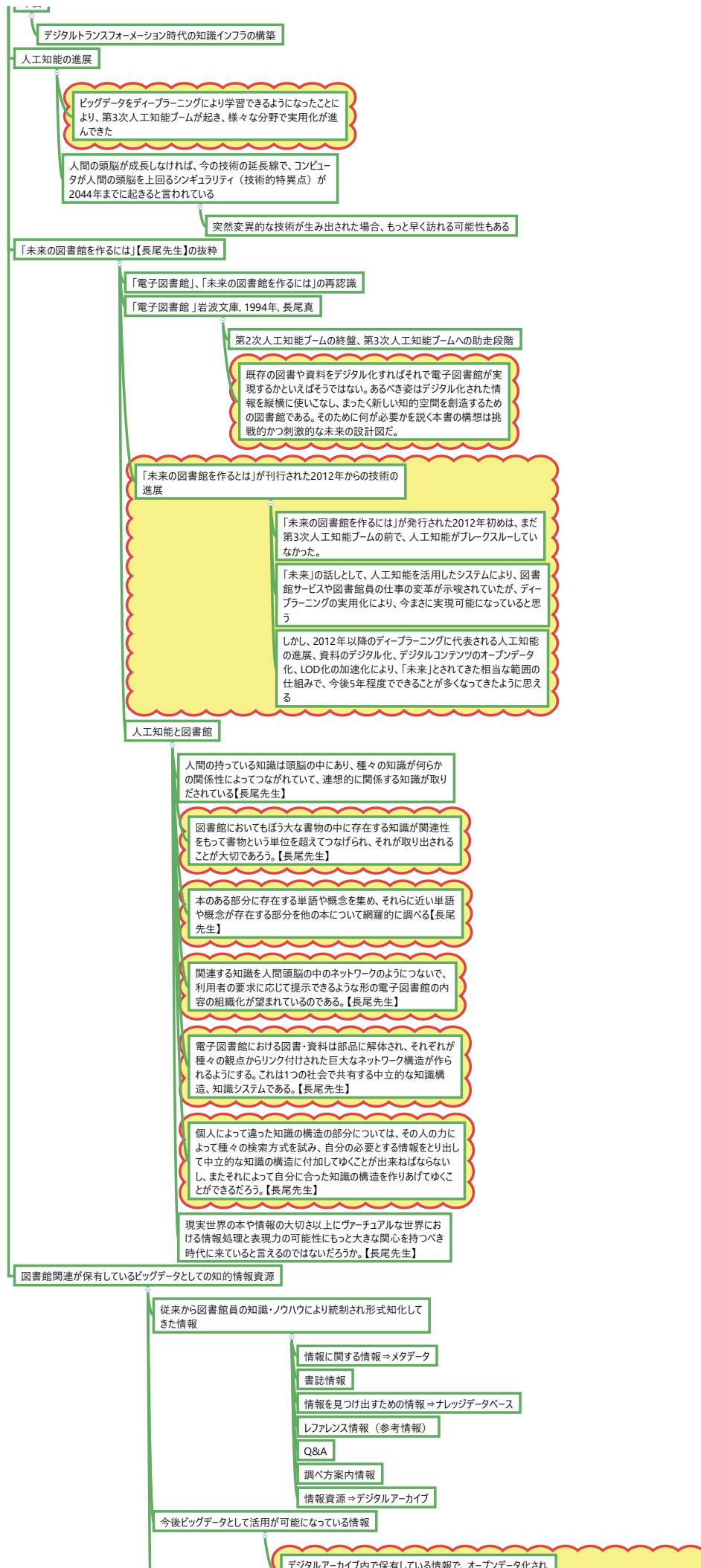
「電子図書館サービス構想」の実現形としての「デジタルコレクション」、「NDLサーチ」、「リサーチナビ」

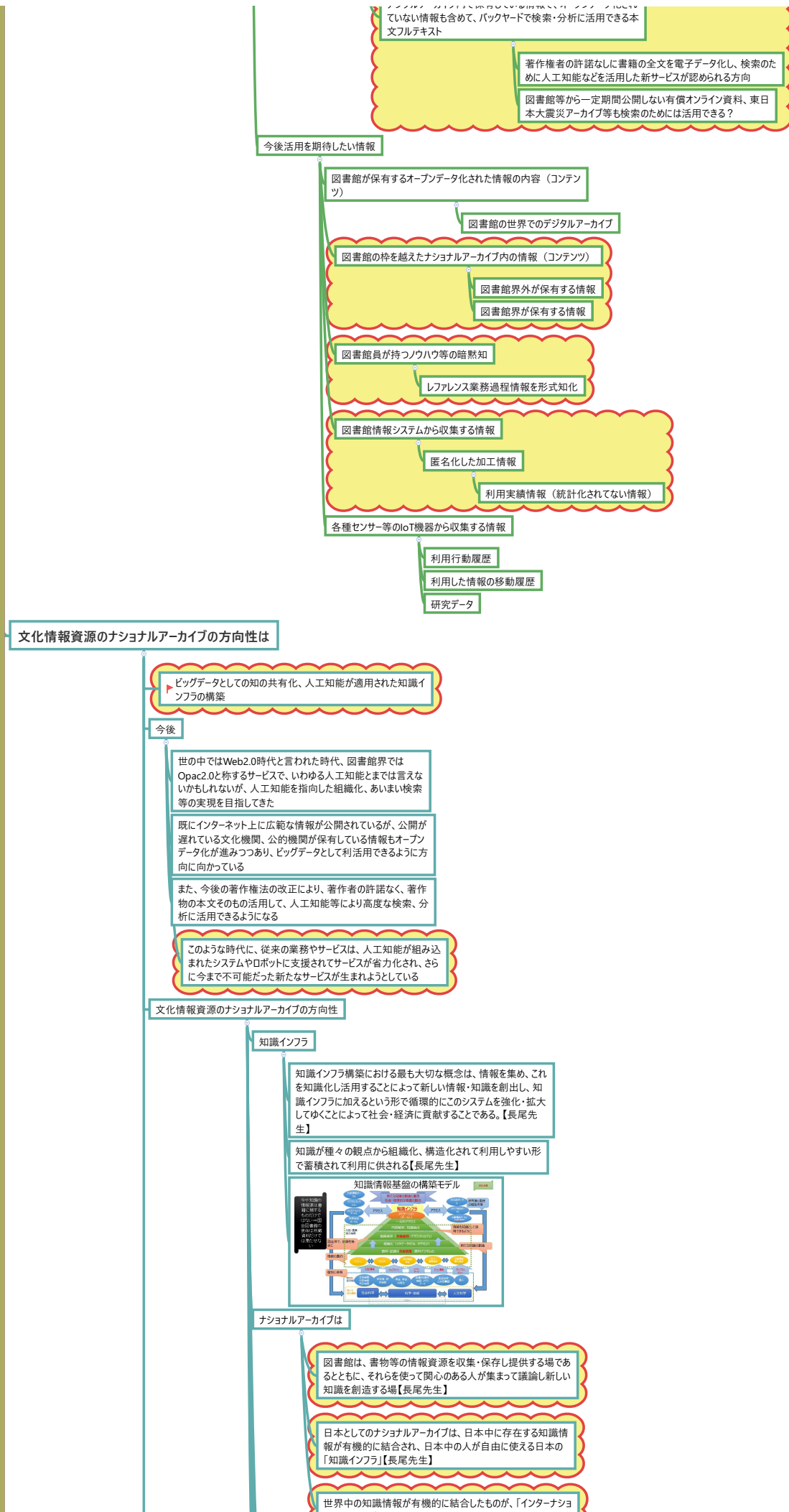
電子図書館サービスの発展形としての「知識インフラ構想」

「知識インフラ構想」の実現形を目指して分野を特定した「東日本大震災アーカイブ」

「東日本大震災アーカイブ」の発展形として分野を文化情報資源まで拡大した「文化資源のナショナルアーカイブ」

#### 今後





## 「ナショナルアーカイブ」

今、ディープラーニングが実用化になった時代に、ナショナルアーカイブはどのような方向へ進んでいくか

図書館での従来から形式知化してきた情報を、信頼性の高い情報として、人工知能の教師用データとして分析し、人工知能として学習した知識として蓄積

書誌、辞書、典拠（シソーラス）類

著作物のイメージ画像、フルテキストを人工知能の学習データとして分析し知識として蓄積

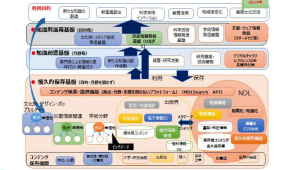
画像、全文フルテキスト（OCR化も含む）、目次、索引、あらすじ、書評、、、

長尾先生が提唱した「電子図書館」が、「人工知能を備えた知識インフラ」となる

文化情報資源としてのナショナルアーカイブがビッグデータとして利用可能になり、人工知能の学習データとして信頼性の高い知識が飛躍的に豊富になり、より精度の高い情報を提供できるようになる

## ナショナルアーカイブの各基盤の概念

## 文化財を含めたナショナルアーカイブの機能イメージ



我が国の多様な文化を知識として保存・継承する役割、様々な分野の専門家が参加し、新たな文化を知識として創造していくための社会的な基盤としての役割、それらの知識を利用目的に応じて発信する役割、そして、これらの仕組みを統括運用していく役割が考えられ、文化的資産を館種毎に集約している拠点と、新たな知識を創造し発信するしている拠点等が分担して構築・運用していくことが必要

## 恒久的保存基盤（あらゆる情報の恒久的保存）

## 恒久保存と利活用のための共通プラットフォーム

1つの機関にすべてを集約するのではなく、各分野のアーカイブを集約する拠点が中核となって分散アーカイブを構築

あたかも1つのアーカイブとして見えるように

個々の情報同士を意味的に関連付け

検索で情報を取り出すだけでなく、取り出された情報から芋づる式に関連する情報を取り出せるように

情報の網羅性・完全性が重要

## 知識創造基盤（新たな知識創造活動の場）

キュレーター、ライブラリアン等の支援のもとで、それぞれの分野の専門家のみならず、広く国民も含めて、情報に付加価値を付けたり、他の分野の情報と関連付けて、二次的著作物を創造する場

## 知識創造支援機能

情報全体の基本情報としてのメタデータを付与する活動、記事、章節項、文節等の単位で組織化・構造化する活動、情報間を意味的に関連付けるための基本情報として、用語辞書、典拠、シソーラス辞書等を作成する活動

## 辞書類作成

## コーパス

人工知能の支援により省力化が進むと思われる

## 件名典拠、シソーラス

本文フルテキストも人工知能への入力データとして特徴量を抽出することにより、書誌データのみならず、本文テキストから標目を抽出することが可能

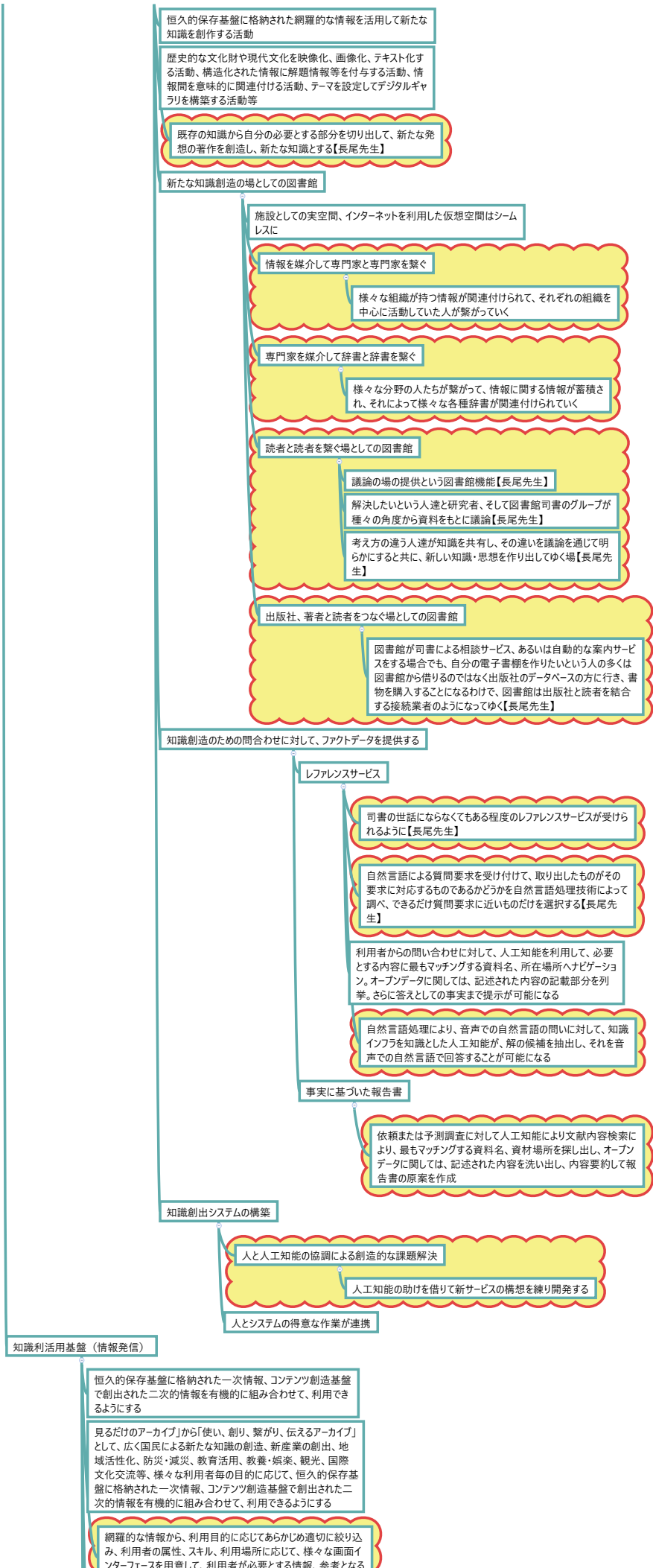
また、コーパスを利用することにより、同義語、上位語、下位語、関連語、分類記号の関係づけをすることも可能ではないか

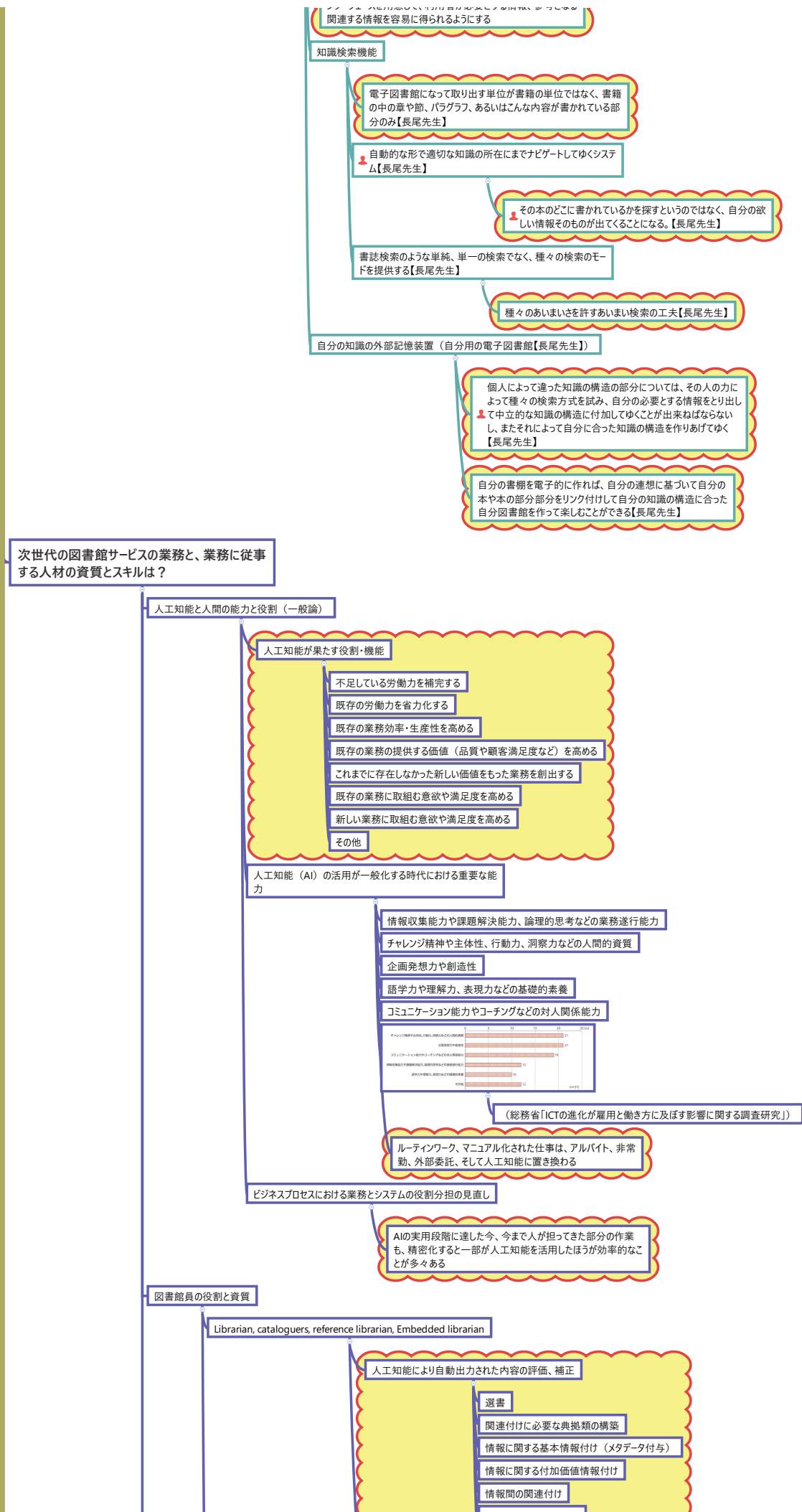
## 固有名典拠

既に刊行されている書籍の著者に関して、人工知能を利用して、著作名、著者、出版年、NDC分類等の特徴量を抽出・分析することにより、著者の自動同定の精度は相当向上すると思われる

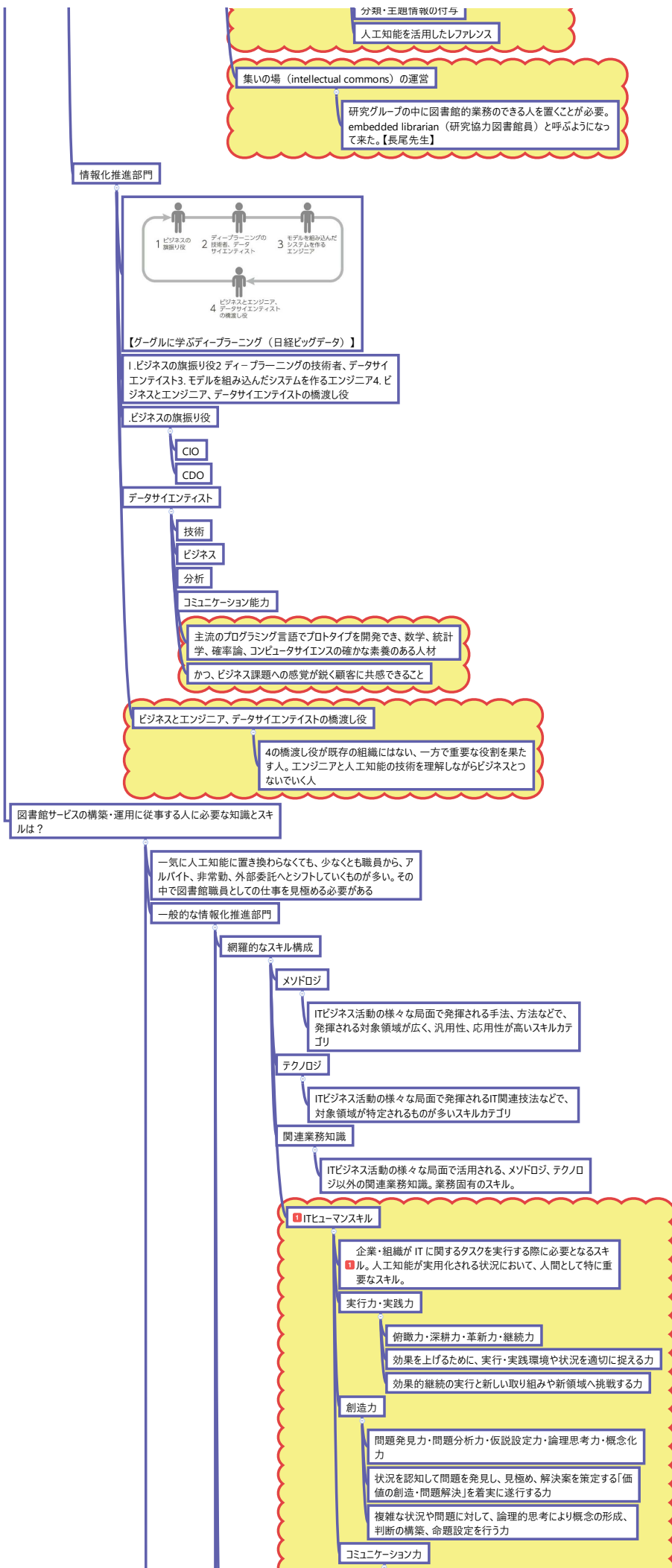
人工知能により、省力化、自動化ができることにより、書籍だけでなく雑誌記事等も含めて、固有名典拠の自動作成が可能になる



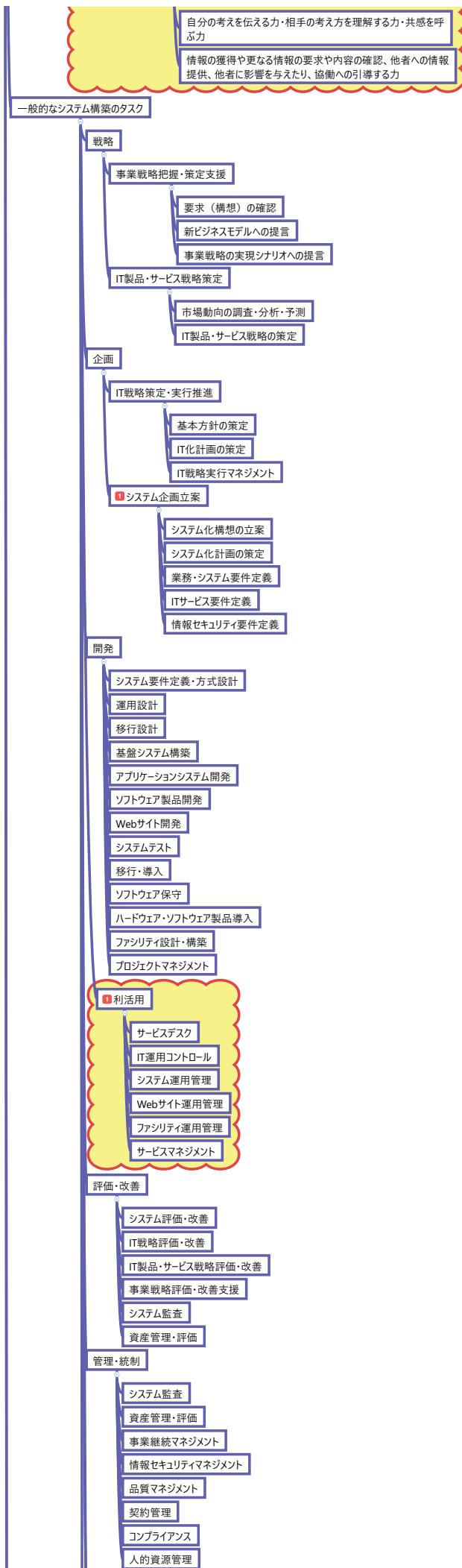


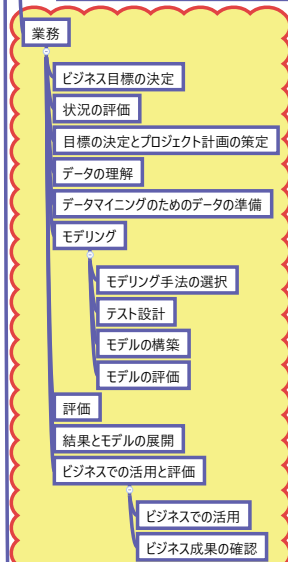
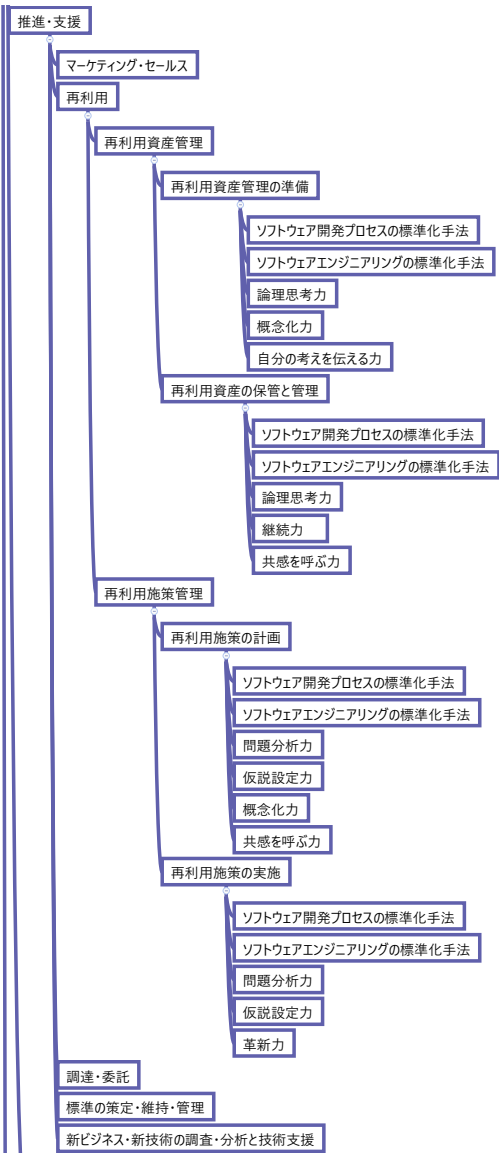












業務タスクとスキル・知識の蓄積のスキーム

