

## 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利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】[リンク](#)

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

匠の技的な高度な伝統的スキルを要する作業や、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と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

デジタルトランスフォーメーション時代の知識インフラの構築

人工知能の進展

ビッグデータをディープラーニングにより学習できるようになったことにより、第3次人工知能ブームが起き、様々な分野で実用化が進んできた

人間の頭脳が成長しなければ、今の技術の延長線で、コンピュータが人間の頭脳を上回るシンギュラリティ（技術的特異点）が2044年までに起きると言われている

突然変異的な技術が生み出された場合、もっと早く訪れる可能性もある

「未来の図書館を作るには」【長尾先生】の抜粋

「電子図書館」、「未来の図書館を作るには」の再認識

「電子図書館」岩波文庫、1994年、長尾真

第2次人工知能ブームの終盤、第3次人工知能ブームへの助走段階

既存の図書や資料をデジタル化すればそれで電子図書館が実現するかといえばそうではない。あるべき姿はデジタル化された情報を縦横に使いこなし、まったく新しい知的空間を創造するための図書館である。そのために何が必要かを説く本書の構想は挑戦的かつ刺激的な未来の設計図だ。

「未来の図書館を作るとは」が刊行された2012年からの技術の進展

「未来の図書館を作るには」が発行された2012年初めは、まだ第3次人工知能ブームの前で、人工知能がブレイクスルーしていなかった。

「未来」の話として、人工知能を活用したシステムにより、図書館サービスや図書館員の仕事の変革が示唆されていたが、ディープラーニングの実用化により、今まさに実現可能になっていると思う

しかし、2012年以降のディープラーニングに代表される人工知能の進展、資料のデジタル化、デジタルコンテンツのオープンデータ化、LOD化の加速化により、「未来」とされてきた相当な範囲の仕組みで、今後5年程度でできることが多くなってきたように思える

#### 人工知能と図書館

人間の持っている知識は頭脳の中にあり、種々の知識が何らかの関係性によってつながれていて、連想的に関係する知識が取りだされている【長尾先生】

図書館においてもぼう大な書物の中に存在する知識が関連性をもって書物という単位を超えてつなげられ、それが取り出されることが大切であろう。【長尾先生】

本のある部分に存在する単語や概念を集め、それらに近い単語や概念が存在する部分を他の本について網羅的に調べる【長尾先生】

関連する知識を人間頭脳の中のネットワークのようにつないで、利用者の要求に応じて提示できるような形の電子図書館の内容の組織化が望まれているのである。【長尾先生】

電子図書館における図書・資料は部品に解体され、それぞれが種々の観点からリンク付けされた巨大なネットワーク構造が作られるようにする。これは1つの社会で共有する中立的な知識構造、知識システムである。【長尾先生】

個人によって違った知識の構造の部分については、その人の力によって種々の検索方式を試み、自分の必要とする情報をとり出して中立的な知識の構造に付加してゆくことが出来ねばならないし、またそれによって自分に合った知識の構造を作りあげてゆくことができるだろう。【長尾先生】

現実世界の本や情報の大切さ以上にヴァーチャルな世界における情報処理と表現力の可能性にもっと大きな関心を持つべき時代に来ていると言えるのではないだろうか。【長尾先生】

# 図書館関係が保有しているビッグデータとしての知的情報資源

従来から図書館員の知識・ノウハウにより統制され形式知化してきた情報

情報に関する情報⇒メタデータ

書誌情報

情報を見つけ出すための情報⇒ナレッジデータベース

レファレンス情報（参考情報）

Q&A

調べ方案内情報

情報資源⇒デジタルアーカイブ

今後ビッグデータとして活用が可能になっている情報

デジタルアーカイブ内で保有している情報で、オープンデータ化されていない情報も含めて、バックヤードで検索・分析に活用できる本文フルテキスト

著作権者の許諾なしに書籍の全文を電子データ化し、検索のために人工知能などを活用した新サービスが認められる方向

図書館等から一定期間公開しない有償オンライン資料、東日本大震災アーカイブ等も検索のためには活用できる？

今後活用を期待したい情報

図書館が保有するオープンデータ化された情報の内容（コンテンツ）

図書館の世界でのデジタルアーカイブ

図書館の枠を越えたナショナルアーカイブ内の情報（コンテンツ）

図書館界外が保有する情報

図書館界が保有する情報

図書館員が持つノウハウ等の暗黙知

レファレンス業務過程情報を形式知化

## 文化情報資源のナショナルアーカイブの方向性は

ビッグデータとしての知の共有化、人工知能が適用された知識インフラの構築

今後

世の中ではWeb2.0時代と言われた時代、図書館界ではOpac2.0と称するサービスで、いわゆる人工知能とまでは言えないかもしれないが、人工知能を指向した組織化、あいまい検索等の実現を目指してきた

既にインターネット上に広範な情報が公開されているが、公開が遅れている文化機関、公的機関が保有している情報もオープンデータ化が進みつつあり、ビッグデータとして利活用できるように方向に向かっている

また、今後の著作権法の改正により、作者の許諾なく、著作物の本文そのものの活用して、人工知能等により高度な検索、分析に活用できるようになる

このような時代に、従来の業務やサービスは、人工知能が組み込まれたシステムやロボットに支援されてサービスが省力化され、さらに今まで不可能だった新たなサービスが生まれようとしている

文化情報資源のナショナルアーカイブの方向性

知識インフラ

知識インフラ構築における最も大切な概念は、情報を集め、これを知識化し、活用する。これによって新しい情報・知識を創出し、知

図書館情報システムから収集する情報

匿名化した加工情報

利用実績情報（統計化されてない情報）

各種センサー等のIoT機器から収集する情報

利用行動履歴

利用した情報の移動履歴

研究データ

知識インフラに加えるという形で循環的にこのシステムを強化・拡大してゆくことによって社会・経済に貢献することである。【長尾先生】

知識が種々の観点から組織化、構造化されて利用しやすい形で蓄積されて利用に供される【長尾先生】



ナショナルアーカイブは

図書館は、書物等の情報資源を収集・保存し提供する場であるとともに、それらを使って関心のある人が集まって議論し新しい知識を創造する場【長尾先生】

日本としてのナショナルアーカイブは、日本中に存在する知識情報が有機的に結合され、日本中の人々が自由に使える日本の「知識インフラ」【長尾先生】

世界中の知識情報が有機的に結合したものが、「インター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

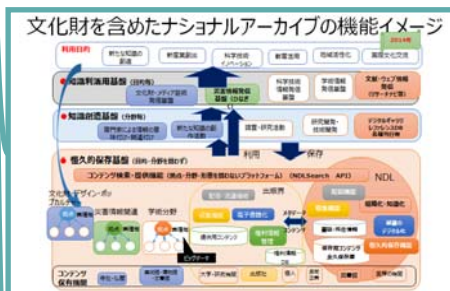
画像、全文フルテキスト（OCR化も含む）、目次、索引、あらすじ、書評、...



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



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し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

知識創生基盤（新たな知識創生基盤の基盤）



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

##### NDC分類

件名典拠をベースにした分類の番号体系は、人間の判断によるところが大きいので、容易に人工知能化することは困難か？

分類を番号体系でなく意味的なシソーラス体系として維持し、改訂の際に体系的に整理し直すことが可能

訂していく作業は、人工知能により体系を整理することは可能。  
また番号体系の版毎の対比表の維持も可能ではないか

ただ、配架のための分類という考え方、資料検索のための分類  
体系は、全文検索を含めて様々なレベルでの検索が容易になっ  
てくる状況において、役割を問い直す必要があるのではないか

情報を関連付けるための情報

共通語彙

分野を超えた情報交換を行うには、個々の単語について、表  
記・意味・データ構造を統一し、互いに意味が通じるようにするこ  
とが必要。人工知能の知識として蓄積する場合、共通語彙にな  
っていることが学習の効率を高められる。

情報と情報を結びつけるための情報

コーパス

国立国語研究所で構築したコーパス

オントロジー

用語辞書

DBPedia等

分野別辞書

大漢和辞書等

分野別シソーラス

JSTシソーラスmap等

個別情報の永続的識別子（Linked Data化）

組織化

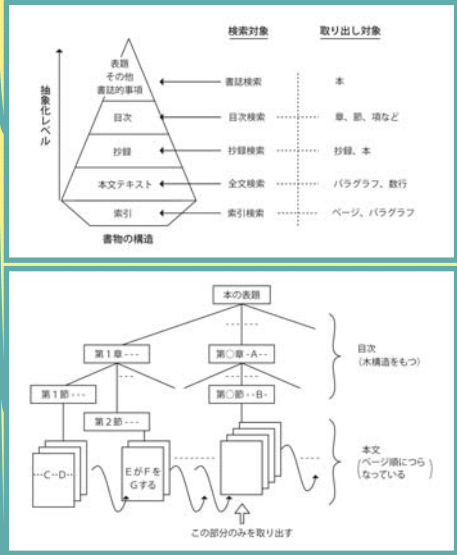
情報に関する情報の組織化（書誌作成）

新しい著作物に対するNDC分類、件名などは、既存の書誌、典  
拠を教師データとして、人工知能により自動付与。専門家は、人  
工知能による精度に応じてチェック、補正することが役割になるの  
ではないか。

\_\_\_\_\_

雑誌記事に関しても、記事を著作単位とした書誌作成が人工知能により省力化して付与が可能になる

図書・資料は部品に解体され、それぞれが種々の観点からリンク付けされた巨大なネットワーク構造が作られるようにする【長尾先生】



目次情報を付けたり、本の表紙の画像や数行の簡単な要旨を付ける

---

従来の書誌情報の考え方をマルチメディア情報に対応できるように拡張かつ詳細化するとともに、対象資料に関係する様々な種類の情報にリンクをはり、それらの情報をたどってゆけるようにする【長尾先生】

全文テキストを利用することにより、新たに時間軸、空間軸に沿った知識の関連付け等の組織化も可能ではないか

レファレンス事例、司書の知識・ノウハウを、信頼性の高い情報を、人工知能の教師用データとして特徴量を自動抽出し、情報を知識化して人工知能の知識データベースとして蓄積

## 知識創造活動

恒久的保存基盤に格納された網羅的な情報を活用して新たな知識を創作する活動

歴史的な文化財や現代文化を映像化、画像化、テキスト化する活動、構造化された情報に解題情報等を付与する活動、情報間を意味的に関連付ける活動、テーマを設定してデジタルギャラリーを構築する活動等

既存の知識から自分の必要とする部分を切り出して、新たな発想の著作を創造し、新たな知識とする【長尾先生】

## 新たな知識創造の場としての図書館

施設としての実空間、インターネットを利用した仮想空間はシームレスに

情報を媒介して専門家と専門家を繋ぐ

様々な組織が持つ情報が関連付けられて、それぞれの組織を中心に活動していた人が繋がっていく

専門家を媒介して辞書と辞書を繋ぐ

様々な分野の人たちが繋がって、情報に関する情報が蓄積され、それによって様々な各種辞書が関連付けられていく

読者と読者を繋ぐ場としての図書館

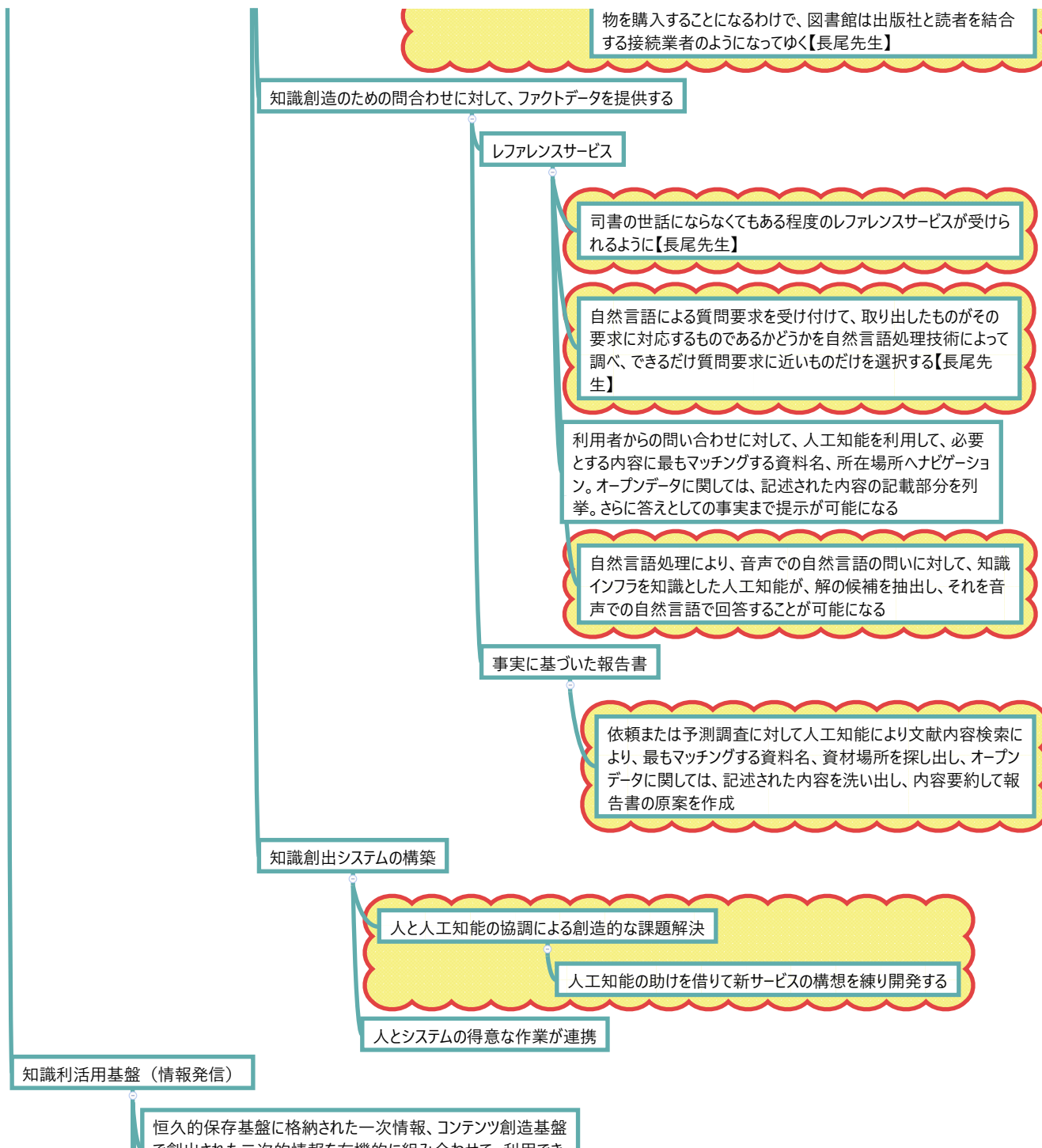
議論の場の提供という図書館機能【長尾先生】

解決したいという人達と研究者、そして図書館司書のグループが種々の角度から資料をもとに議論【長尾先生】

考え方の違う人達が知識を共有し、その違いを議論を通じて明らかにすると共に、新しい知識・思想を作り出してゆく場【長尾先生】

出版社、著者と読者をつなぐ場としての図書館

図書館が司書による相談サービス、あるいは自動的な案内サービスをする場合でも、自分の電子書棚を作りたいという人の多くは図書館から借りるのではなく出版社のデータベースの方に行き、書



で抽出された一次的情報を有機的に組み合わせ、利用できるようにする

見るだけのアーカイブから「使い、創り、繋がり、伝えるアーカイブ」として、広く国民による新たな知識の創造、新産業の創出、地域活性化、防災・減災、教育活用、教養・娯楽、観光、国際文化交流等、様々な利用者毎の目的に応じて、恒久的保存基盤に格納された一次情報、コンテンツ創造基盤で創出された二次的情報を有機的に組み合わせて、利用できるようにする

網羅的な情報から、利用目的に応じてあらかじめ適切に絞り込み、利用者の属性、スキル、利用場所に応じて、様々な画面インターフェースを用意して、利用者が必要とする情報、参考となる関連する情報を容易に得られるようにする

#### 知識検索機能

電子図書館になって取り出す単位が書籍の単位ではなく、書籍の中の章や節、パラグラフ、あるいはこんな内容が書かれている部分のみ【長尾先生】

自動的な形で適切な知識の所在にまでナビゲートしてゆくシステム【長尾先生】

その本のどこに書かれているかを探すというのではなく、自分の欲しい情報そのものが出てくることになる。【長尾先生】

書誌検索のような単純、単一の検索でなく、種々の検索のモードを提供する【長尾先生】

種々のあいまいさを許すあいまい検索の工夫【長尾先生】

#### 自分の知識の外部記憶装置（自分用の電子図書館【長尾先生】）

個人によって違った知識の構造の部分については、その人の力によって種々の検索方式を試み、自分の必要とする情報をとり出して中立的な知識の構造に付加してゆくことが出来ねばならないし、またそれによって自分に合った知識の構造を作りあげてゆく【長尾先生】

自分の書棚を電子的に作れば、自分の連想に基づいて自分の本や本の部分部分をリンク付けて自分の知識の構造に合った自分図書館を作って楽しむことができる【長尾先生】

# 次世代の図書館サービスの業務と、業務に従事する人材の資質とスキルは？

## 人工知能と人間の能力と役割（一般論）

### 人工知能が果たす役割・機能

不足している労働力を補完する

既存の労働力を省力化する

既存の業務効率・生産性を高める

既存の業務の提供する価値（品質や顧客満足度など）を高める

これまでに存在しなかった新しい価値をもった業務を創出する

既存の業務に取り組む意欲や満足度を高める

新しい業務に取り組む意欲や満足度を高める

その他

## 人工知能（AI）の活用が一般化する時代における重要な能力

情報収集能力や課題解決能力、論理的思考などの業務遂行能力

チャレンジ精神や主体性、行動力、洞察力などの人間的資質

企画発想力や創造性

語学力や理解力、表現力などの基礎的素養

コミュニケーション能力やコーチングなどの対人関係能力



（総務省「ICTの進化が雇用と働き方に及ぼす影響に関する調査研究」）

ルーティンワーク、マニュアル化された仕事は、アルバイト、非常勤、外部委託、そして人工知能に置き換わる

## ビジネスプロセスにおける業務とシステムの役割分担の見直し



AIの実用段階に達した今、今まで人が担ってきた部分の作業も、精密化すると一部が人工知能を活用したほうが効率的なことが多々ある

## 図書館員の役割と資質

Librarian, cataloguers, reference librarian, Embedded librarian

人工知能により自動出力された内容の評価、補正

選書

関連付けに必要な典拠類の構築

情報に関する基本情報付け（メタデータ付与）

情報に関する付加価値情報付け

情報間の関連付け

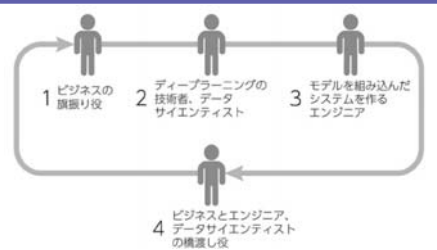
分類・主題情報の付与

人工知能を活用したレファレンス

集いの場（intellectual commons）の運営

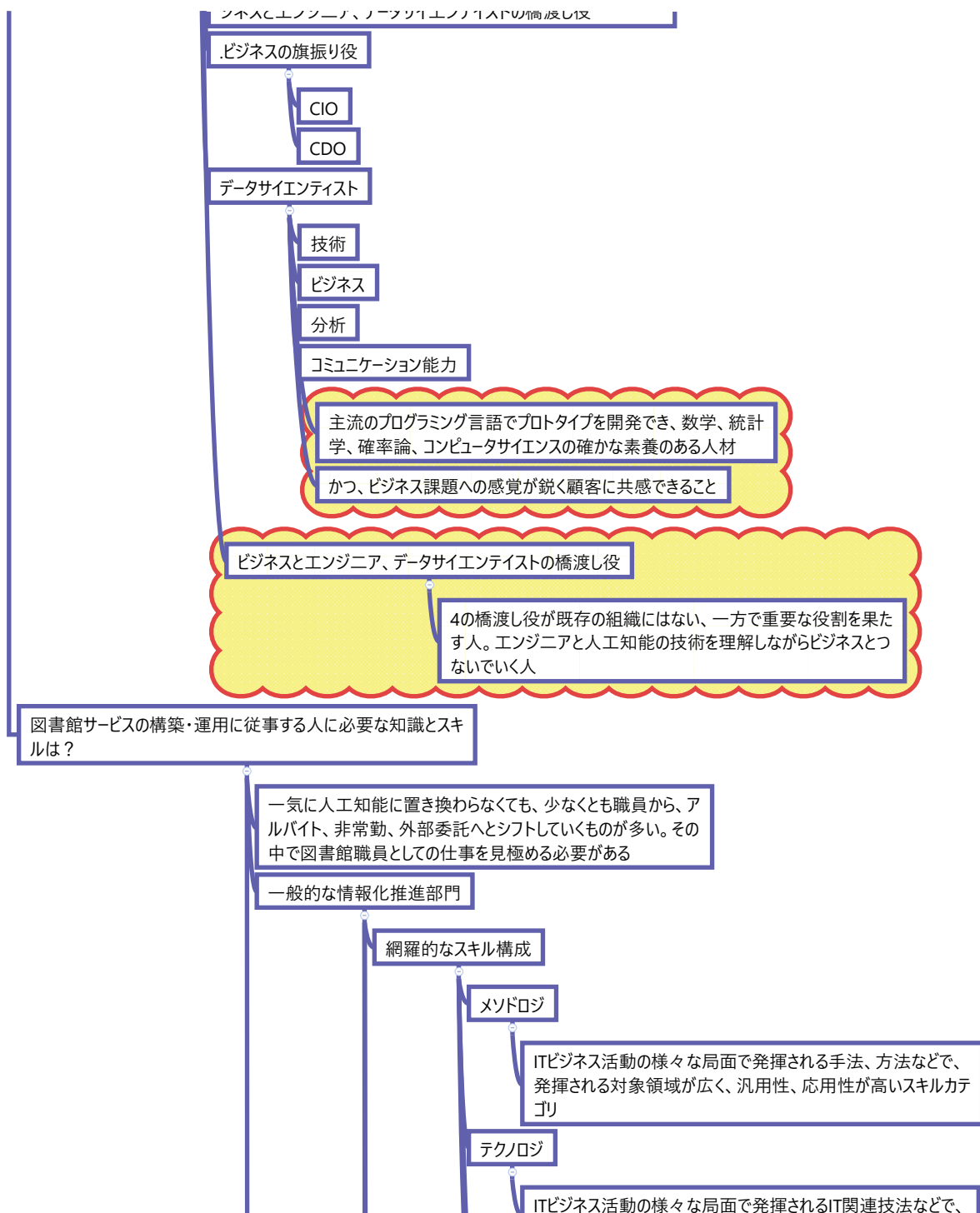
研究グループの中に図書館的業務のできる人を置くことが必要。  
embedded librarian（研究協力図書館員）と呼ぶようになって来た。【長尾先生】

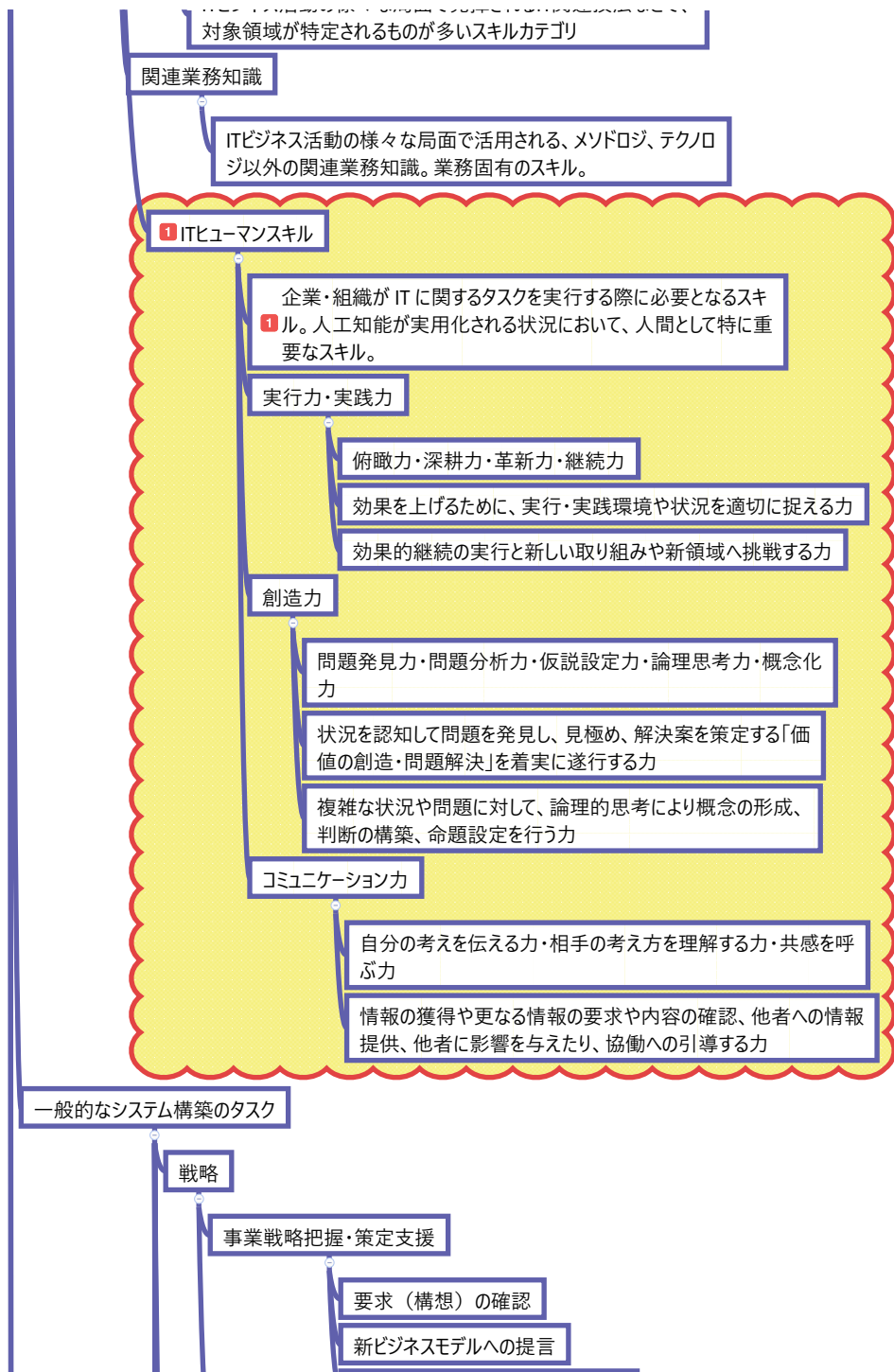
## 情報化推進部門

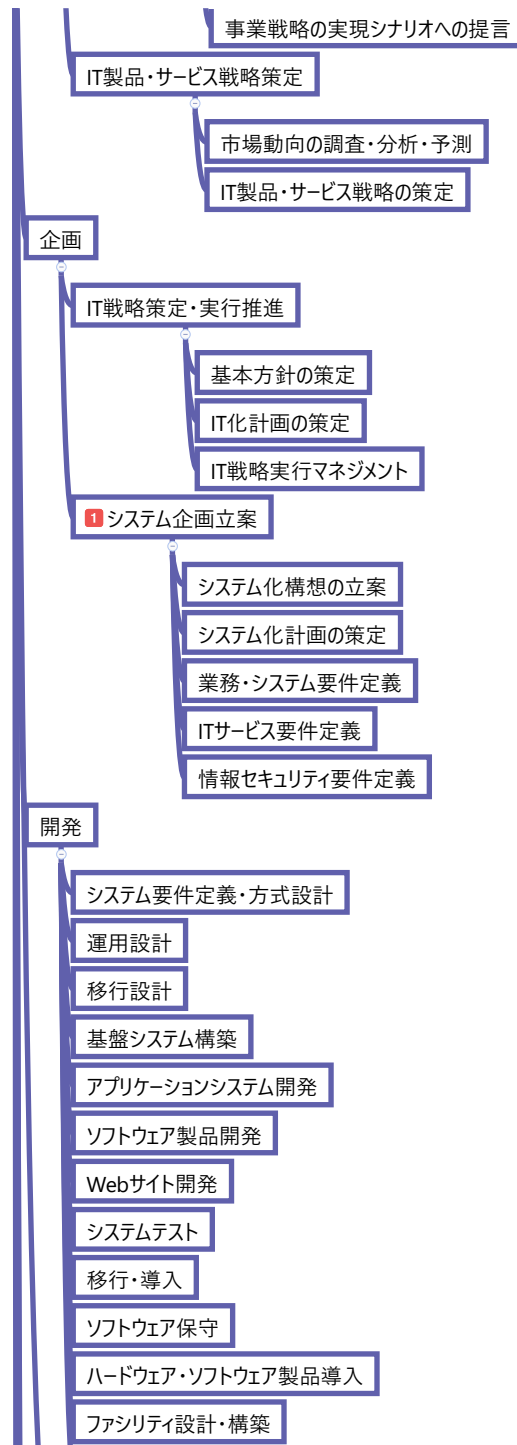


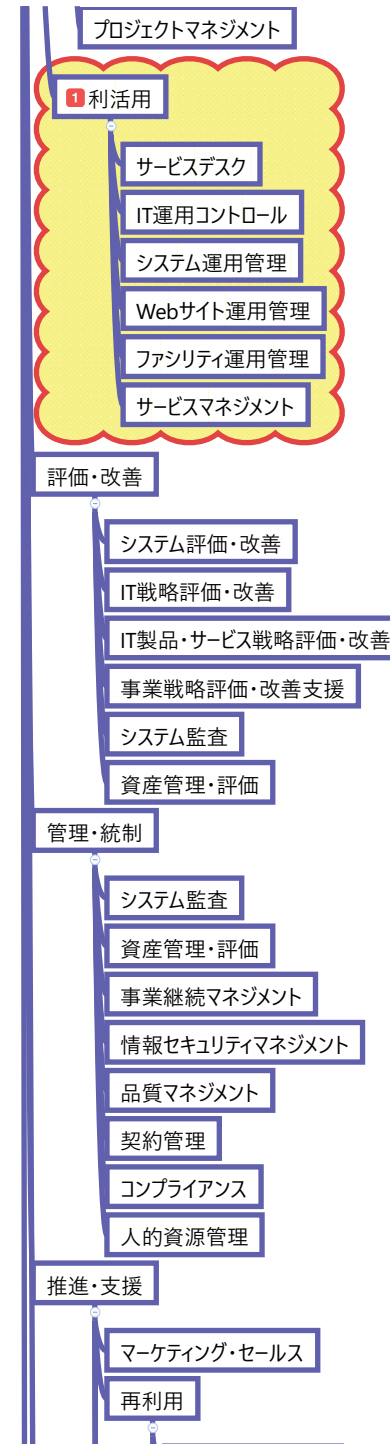
【グーグルに学ぶディープラーニング（日経ビッグデータ）】

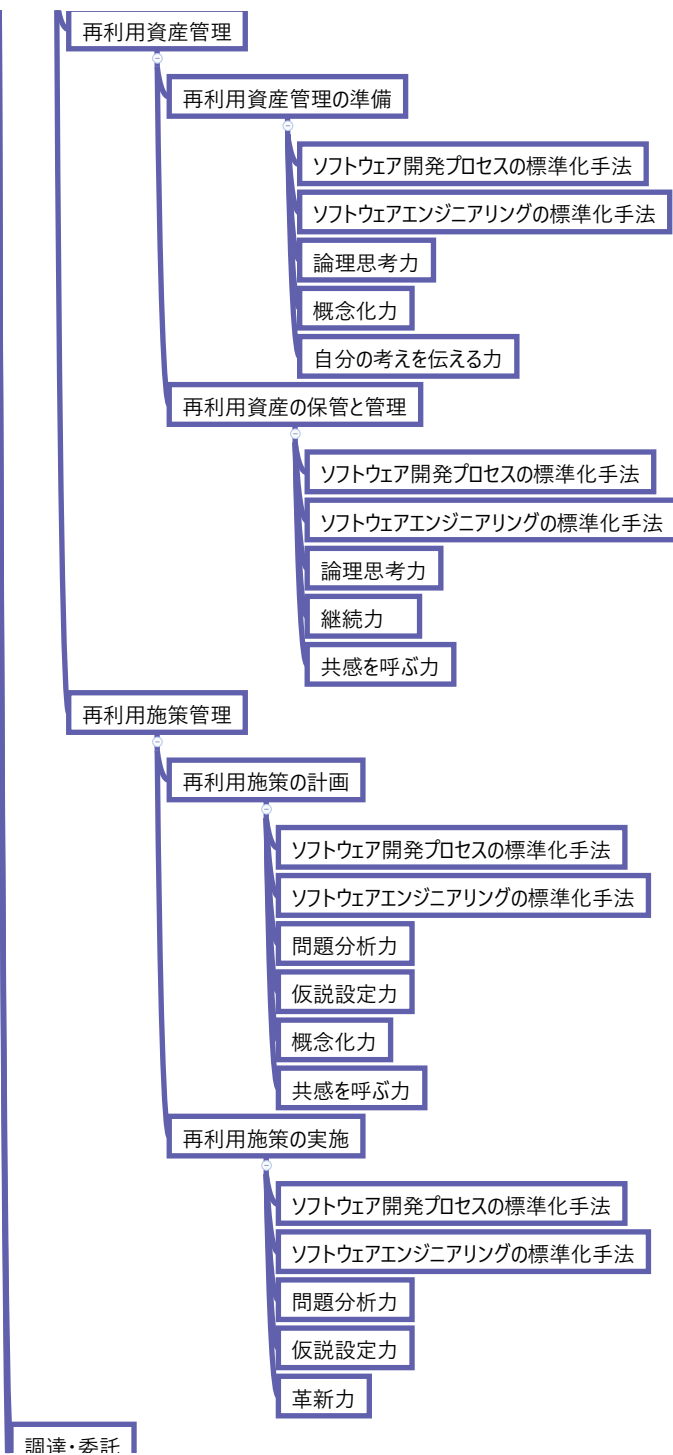
1. ビジネスの旗振り役 2. ディープラーニングの技術者、データサイエンティスト 3. モデルを組み込んだシステムを作るエンジニア 4. ビジネスとエンジニア、データサイエンティストの構築し役











調達・委託

標準の策定・維持・管理

新ビジネス・新技術の調査・分析と技術支援

業務

ビジネス目標の決定

状況の評価

目標の決定とプロジェクト計画の策定

データの理解

データマイニングのためのデータの準備

モデリング

モデリング手法の選択

テスト設計

モデルの構築

モデルの評価

評価

結果とモデルの展開

ビジネスでの活用と評価

ビジネスでの活用

ビジネス成果の確認

政府標準ガイドラインに沿った開発タスクとドキュメント



業務タスクとスキル・知識の蓄積のスキーム

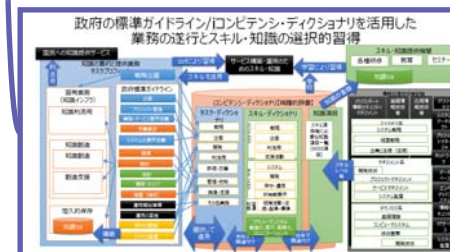
業務遂行のタスクとスキル・知識の蓄積の関係







iコンピテンシ・ディクショナリを活用した業務の遂行とスキル・知識の選択的習得方法



今後の図書館サービスの実現のためのタスクと必要なスキル

文化資源の収集・保存・修復・公開のスキル

図書館が扱う情報資産は、文献だけではない。

専門分野に関する知見（文化・芸術・学術）

文化資源を取り扱うための知識・技能

保存・修復技術

文化資源に価値を見出し、情報として記述するカタログラーとしての知識・技能

文化資源の価値を顕在化させて共有するための企画・発信するキュレーターとしての知識・技能

文化資源と人々をつなぎ、新たな価値を創出するコーディネータ、エンベデッドライブラリアンとしての知識・技能

文化資源を扱う活動の使命を明らかにし、その達成に向け経営資源を配分し、事業を統括するマネージャーとしての知識・技能

デジタル技術を活用したアーカイブ化のための知見

文化資源を取り扱う様々な局面でITを活用し、文化資源をデジタル化し情報メディアに乗せていく技術を有する

著作権をはじめとする知的財産権、肖像権、契約など各種法律分野に関する知識を有する

文化資源を情報として収集・組織化・保存し、公開することを実現するシステムの開発・運用管理の知識・技能

## データサイエンス領域のスキル標準「ITSS+」の適用【要約】

### 位置付け

iコンピテンシ・ディクショナリの補足として活用

従来のiコンピテンシ・ディクショナリでは、まだに辞書化が十分でない領域

★ 知識インフラの構築は、ビッグデータや人工知能の活用が必須であり、典型的なデータサイエンス領域として位置づけられる

### スキルカテゴリ

#### ビジネス

課題背景を理解した上で、ビジネス課題を整理し、解決する。

#### データサイエンス

情報処理、人工知能、統計学などの情報科学系の知恵を理解し、活用する。

#### データエンジニアリング

データサイエンスを意味のある形に使えるようにし、実装、運用する。

効率的・効果的なシステム開発を行うシステムエンジニア

基礎となる学問分野の知識

A：自然言語分野

B：音声・音楽分野

C：画像・映像分野

D：コンピュータ・ソフトウェア、情報通信

E：知識工学、人工知能

F：図書館学、図書館情報学

詳細は別途

文化資源の保有機関にも、先進技術の研究開発および実用化を目指す研究者

