

ナレッジグラフで実現する 「トラストで説明可能なAI」と社会実装

“Trustworthy and Explainable AI” Achieved Through Knowledge Graphs and Social Implementation

富士 秀 中澤 克仁 吉田 宏章

あらまし

人工知能（AI）の性能が向上する一方、実用場面で専門家がAIを利用する際には、AIの答えをそのまま提示するだけでは十分な判断材料が得られず、結果として十分な信頼性が得られないという問題がある。そのため近年では、AIの説明可能性に関する研究開発がグローバルな規模で行われるようになってきている。この問題に対して富士通研究所では、専門知識を体系的に表現することができるナレッジグラフを活用して理由や根拠を付与する、「トラストで説明可能なAI」を世界に先駆けて実現した。これによって、利用者に十分な判断材料を提供し、AIの信頼性を向上させることができる。

本稿ではまず、トラストで説明可能なAIのベースとなっているナレッジグラフについて概観する。次に、金融、化学の各分野における取り組みについて紹介する。

Abstract

While the performance of AI is improving, when experts use AI in practical applications, presenting AI's inferences as is may not provide adequate information for making decisions. This failure to ensure sufficient reliability is a problem. Accordingly, R&D on the Explainability of AI has come to be conducted on a global scale recently. In response to this problem, Fujitsu has achieved the world's first “trustworthy and explainable AI” that makes use of knowledge graphs capable of representing expertise in a systematic manner to explain the reason and basis. This provides users with sufficient information to make decisions and allows the reliability of AI to be improved. This paper first outlines a knowledge graph that provides the base for Fujitsu's trustworthy and explainable AI. It then goes on to focus on use in the industrial sector and describe initiatives in the financial and chemical sectors.

1. まえがき

医療、金融、化学などの高い信頼性を求められる産業分野における人工知能（AI）活用では、AIが導き出した答えを分野の専門家が解釈してから利用することになる。しかし、AIを利用する専門家にとって、単なる「Yes/No」や「90%」などというAIによる推定結果だけでは判断材料としては不十分であり、AIの推定結果を採用するかどうか信頼性を持って判断することができないという問題がある。そのため、近年「説明可能なAI」の研究が幅広く行われるようになってきた。

これまでの説明可能なAIでは、主にディープラーニングを中心とした機械学習に説明機能を加えることによって、説明可能性を向上させる取り組みが中心であった。しかし、この取り組みだけでは産業分野で必要となる信頼性を十分に担保できなかった。

この問題に対して富士通研究所では、ナレッジグラフによってAIの推定結果に理由や根拠を付与することで信頼性を向上させる、「トラストで説明可能なAI」の社会実装を世界に先駆けて実現した。

本稿ではまず、トラストで説明可能なAIのベースとなっているナレッジグラフについて概観する。次に、金融、化学の各分野における取り組みについて紹介する。

2. 説明可能なAIとその課題

本章では、説明可能なAIの概要とその課題について述べる。

2.1 説明可能なAIとは

AIはデータから統計的特徴を学習し、未知の事象を予見することができる。その予見に至ったAIの思考ともいえるべき理由や根拠を説明できるAIが、説明可能なAIである。

現在行われている説明可能なAIの研究開発では、AIの学習および推定に用いるディープラーニングにおいて、入力データのどの統計的特徴がAI推定の決め手になったかを特定する取り組みが主流となっている。例えば、説明可能なAIの一種である

Deep explanationでは、画像データに写ったものが猫かどうかを区別するのに、特徴として用いた耳部分を認識したAIニューロンの活性度を提示することで、AIの思考を表現するなどの方法を取り入れている。なおAIニューロンとは、ディープラーニングで用いるニューラルネットワークを構成するノードのことで、人間の脳で言えば神経細胞に相当する。

このような画像認識の例では、耳部分が決め手になっているということ特定し、それを入力画像上で利用者に提示すれば、利用者は視覚的にAIの判断理由を推察できる。この方向性の研究は現在盛んに進められており、例えば画像に対する認識のように人間が視覚的に判断しやすい分野では、実用レベルに近い成果も出てきている。

2.2 説明可能なAIの課題

前節で述べたように、AIの判断理由を示すことによって一定の説明可能性の向上を実現できる。しかし、これだけでは産業分野などの実用場面で求められる信頼性を担保できていないのが実情である。信頼性を担保するためには、AIによる推定結果に対して、分野専門家が参照できる根拠を併せて提示することが望まれる。

例えば医療分野において、大量の患者データとその診断結果を学習データとして学習させたAIに、診断の推定をさせるシステムを想定する。この場合、利用する医師に対して、どの患者データがAIの判断理由になったかを示すだけでは不十分である。その判断に関連の深い根拠となる論文などを併せて提示することが、そのAIによる推定結果を専門家が利用する上で重要になる。

3. トラストで説明可能なAI

富士通研究所は今般、ナレッジグラフを使ってAIによる推定結果に根拠を付与することで、AIの信頼性を向上させるトラストで説明可能なAIを世界に先駆けて実現した。このうちナレッジグラフは、専門家の知識が蓄積された文献情報を用いて構築した。

ナレッジグラフによる説明可能なAIの構成を、

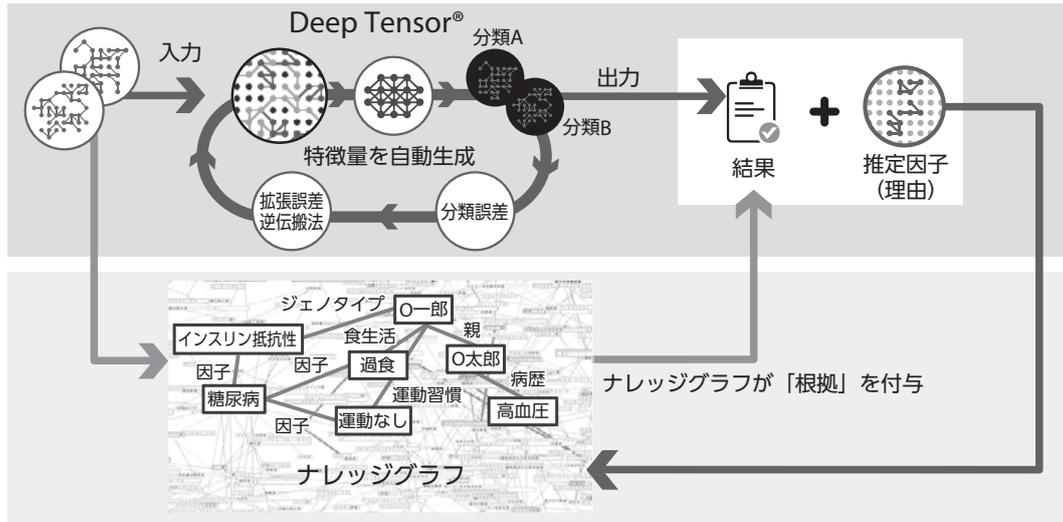


図-1 ナレッジグラフによる説明可能なAIの構成

図-1に示す。左上からの入力に対して、グラフ構造のデータを学習するDeep Tensorが推定結果および推定因子（理由）を出力し、これを基に下半分のナレッジグラフが結果に対する根拠を付与する構成となっている。

ナレッジグラフとは、専門分野などの知識を体系的に表現するためのフレームワークである。文献などの専門知識を基にして、AIによる推定結果に根拠を付与するためには、知識をコンピュータが扱える形式しておく必要がある。そのため、本研究ではナレッジグラフを用いることで知識をコンピュータ上で表現した。

知識の表現形式にはいくつかの手法がある。その一つに、概念間のつながりとして表現する方法があり、ナレッジグラフはその代表例である。概念間のつながりとして表現する方法はシンプルな構造となる傾向があり、AIなどの機械処理に向いていると考えられる。

ナレッジグラフの構成を、図-2に示す。下半分に、概念および概念間のつながりによる知識体系が表され、上半分にこれらの概念間の関係性に関する分野固有の定義が概念化の明示的な仕様であるオントロジーとして記載されている。例えば、過食は糖尿病の原因の一つであるという知識は、「過食」と「糖尿病」という二つの概念の間に「原因」というつながりがあると表現する。ナレッジグラフにメタデー

タやオントロジーなどを用いて意味付けを行うことで、AIは概念間のつながりの意味をより深く捉えられるようになる。我々人類が持つ知識をナレッジグラフに移転していくことで、一人ひとりの人間が扱える知識を越えて、複数人の知識が詰まった知識を構成できる。

なお、図の上半分に示す分野固有の定義は、その分野の「専門家」によって記述されることを表す。また、下半分に示す概念および概念のつながりは、論文やウェブ情報などの一般に公開された「公開データ」から作成する部分と、企業や大学などの組織が内部に保有し公開されることのない「非公開データ」から構成される。ここで、公開データにおけるLOD (Linked Open Data) はウェブ上でデータを公開して共有するための方法である。

4. 金融分野への適用事例

富士通研究所では、今回開発したトラストで説明可能なAIの技術を、ゲノム医療、金融、化学などの、高い信頼性が求められる実際の産業分野への適用を進めようとしている。本章では、金融分野への適用事例について述べる。なお、ゲノム医療分野への適用⁽¹⁾については、別稿⁽²⁾で解説している。

近年、金融分野ではICTを駆使した金融商品・サービスであるFinTech（フィンテック）への取り

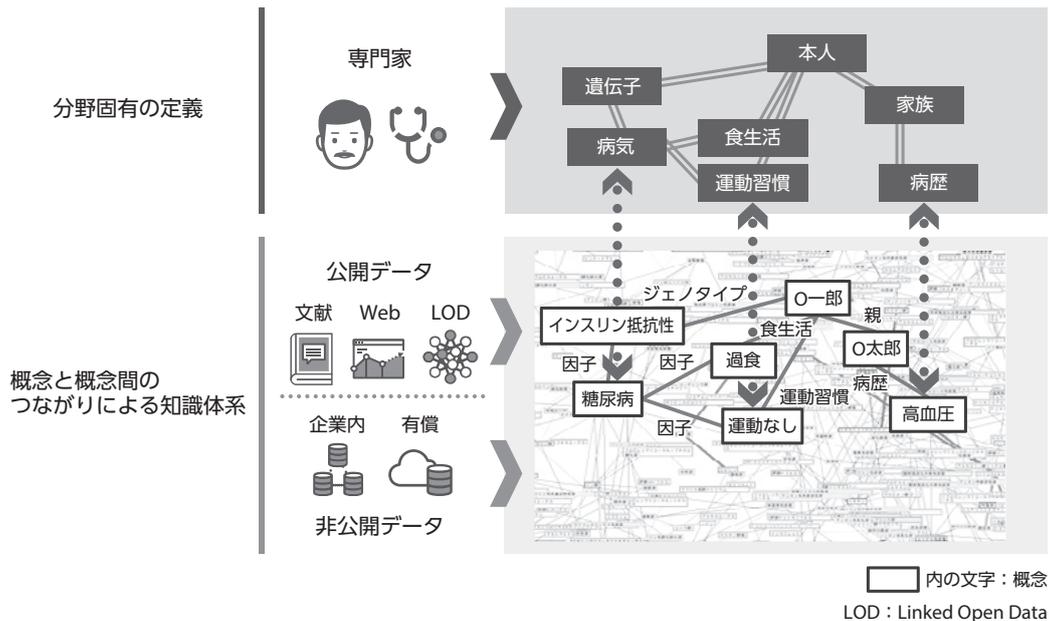


図-2 ナレッジグラフの構成

組みが積極的に進められている。特に銀行の融資業務においては、信用度スコアにAIを導入するなど、融資審査の自動化・効率化が検討されている。⁽³⁾

一方で、このミッションクリティカルな金融分野にAIを導入するためには、AIから出力された結果に対する高い納得性が得られる説明が求められる。この課題を完全に解決できていないことが、AI導入の大きな障害となっている。

現在、多くのAI融資審査では、結果に影響した特徴量を特定する方法によってAIの出力結果に至った理由を説明するものが主流である。しかし、銀行が融資先から、より納得性の高い説明を求められた際には、AIがその特徴量を重要と判断した根拠について、人間が理解できるレベルまで掘り下げたトラストで説明可能なAIが必要になる。

富士通研究所では、金融分野にトラストで説明可能なAIを適用することによって、銀行が保有する企業間の為替取引データをグラフ構造に変換して入力データとし、信用スコア（出力結果）から成長企業（銀行が定める融資判定のしきい値より上位か下位か）を判定する際に、トラストで説明可能なAIの適用を目指した。それを実現するために、Deep Tensorから出力された推定因子（出力結果に寄与した部分グラフ）を基に、金融ナレッジグラフを用

いることで、Deep Tensorが出力した結果に至った根拠を説明する技術を開発した。

金融ナレッジグラフとは、国税庁から公開されている法人番号を名寄せ（データの中から同じ人物、企業などをまとめる作業）する際の基本情報として、金融庁のEDINETで公開されている有価証券報告書を中心としたオープンデータと、銀行が独自に保有している企業情報（企業間の為替取引データなど）とを突き合わせて構成した、グラフ形式の知識ベースである。^{(4), (5)} これによって、ミッションクリティカルな金融分野においても、AIによる高精度な融資判定結果が出力されるだけでなく、この結果に起因した推定因子に対する明確な根拠も説明できるトラストで説明可能なAIを提供できるようになる。

金融ナレッジグラフから融資判定の根拠を分析するツールを、図-3に示す。この分析ツールは、大きく分けて三つのパネル情報で構成されている。左上には、Deep Tensorから出力された推定因子を出力結果に対して寄与率（影響度）が高い順に並べた予測寄与度一覧、左下には、推定因子間の具体的な企業間取引情報（取引量と取引回数の時系列データ）、右には、推定因子間の関係性（取引、役員、関係会社、大株主）がグラフネットワーク形式で表示される。

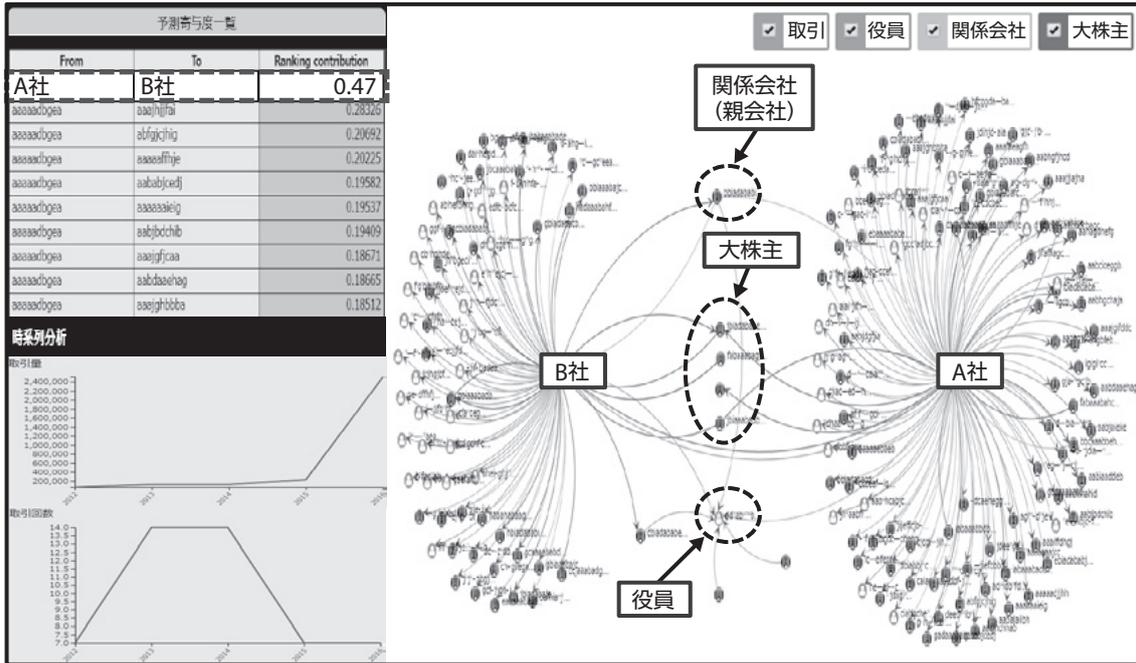


図-3 金融ナレッジグラフの分析画面

このツールを使って、A社の信用リスクを分析したケースを紹介する。図-3の事例では、まず左上のリストから、Deep Tensorから出力された推定因子の中で、A社からB社への為替取引が判定結果に与えた影響が最も大きいことが分かる。次に、時系列分析に示されているA社からB社への経年的な取引データを分析することで、取引量が近年著しく増加している状況や、取引回数が確保されている状態を観察できる。これによって、Deep Tensorから出力された結果の妥当性を確認できる。

更に、企業間の関係性を可視化したグラフネットワークを用いることで、Deep Tensorが重要と判定した推定因子について、より深掘りした背景知識を与えてくれる。例えば、A社の信用スコアにおいて、Deep Tensorが最も重要な関係と判断したA社とB社の間には、共通の役員が存在している上に、大株主も多数存在しており、更に親会社が共通しているという関係性があることが分かる。これらの状況を考慮すると、A社とB社の関係性は非常に強固である上に、共通役員や株保有関係、同じ親会社の傘下にあることで企業活動の安定性が確保されていることが推測できる。よって、銀行の融資先としてリスクが低い企業であることが推察できる。

このように、金融分野へのトラストで説明可能なAIの提供に向けて、オープンデータなどの外部情報で構築された金融ナレッジグラフを活用することで、Deep Tensorが重要と判断した推定因子の根拠に関する説明が可能となる。更に、この説明能力の向上に伴い、将来的にAIの信頼性が高まり、金融分野へ広く展開されることが期待される。

5. 化学分野への適用事例

材料の多くは、分子単体で機能発現することは少なく、集合体の高次構造が機能を左右する。また、多くの有機材料は複合材料や組成物の状態で利用される。そのため、材料の設計や解析においては、構成要素や製造プロセスを含めた複雑な関係を理解していく必要がある。富士通研究所では、機械学習などで得た材料設計指針や化学現象に対して根拠の説明を行うことで、材料の設計や解析の信頼性を高める取り組みを進めている。本章では、化学分野でナレッジグラフを適用した事例について紹介する。⁽⁶⁾

富士通研究所で開発している化学ナレッジグラフは、主に化学物質名やその基本物性に関する情報を蓄積した、例えば日本化学物質辞書⁽⁷⁾やPubChem⁽⁸⁾

などの化学物質データベース、化学に関連する特許公報、公開されている製品情報などの関係性を基に構築されている。また、自然言語処理によって論文から抽出した化合物の物性や製造プロセスなど、化学反応に関する様々な事例も情報源とすることで、説明可能なAIの信頼性を向上させている。化学ナレッジグラフの狙いは、幅広いデータ間で、より深く関係性を扱えるという特徴を活かすことである。

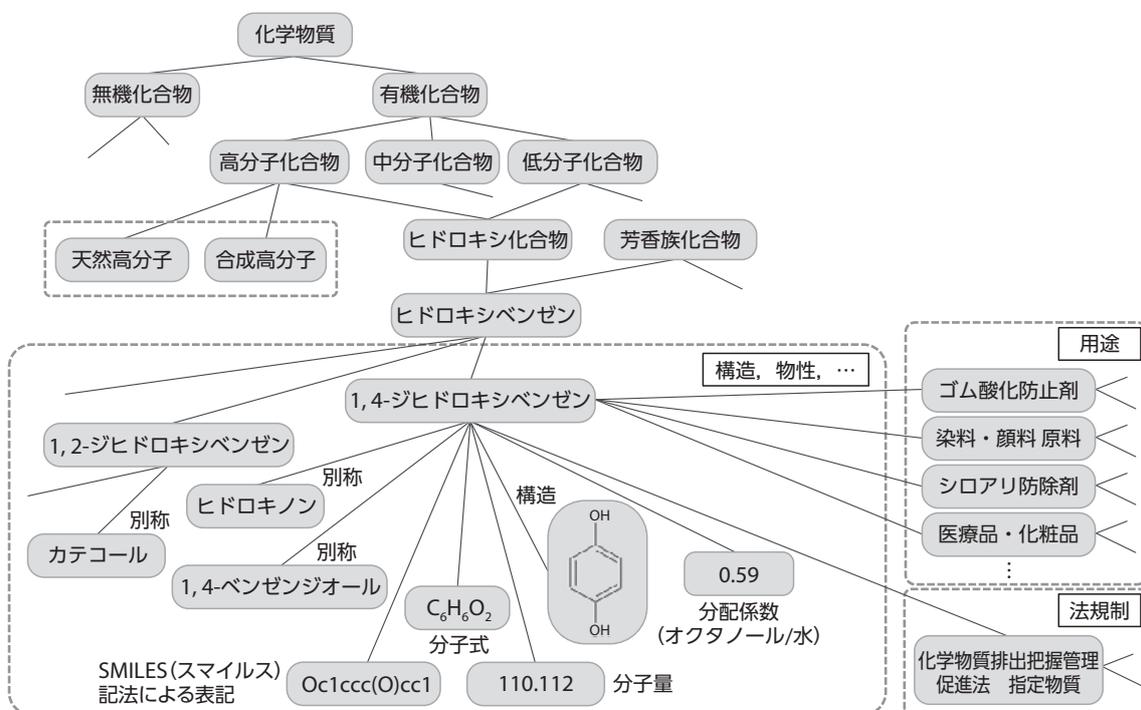
近年の化学分野では、材料に求められる機能が多様化・高度化しており、人の経験と勘に依存した従来の材料開発手法では限界がある。これに代わって、AIを用いて多くのデータから機能発現の因子を見つけ出す、いわゆる材料インフォマティクス的手法に注目が集まっている。一方、多くの研究テーマでは、機械学習による解析に十分な量のデータを用意できない、あるいはシミュレーションや機械学習から導出された材料設計結果の信頼性が低いという問題がある。

この問題に対して富士通研究所は、化学ナレッジグラフを用いて、化合物や物性、組成、用途などの

分野を越えた多様な関係性、類似性から少量のデータを補完し、材料の設計や解析の説明力を向上させることに取り組んでいる。単なる分子構造と物性の相関だけではなく、製造プロセスや反応メカニズムの関係性の特徴を加味することで、材料開発の信頼性を向上させる。

化学ナレッジグラフの構造例を図-4に示す。化学物質同士の関係を表すには、例えば、同種の官能基や同様の性質を有する化合物群をオントロジー的にひも付けていく。例えば、1,4-ジヒドロキシベンゼンという化合物は、ベンゼンを基本骨格としてヒドロキシ基が置換した化合物群に対する総称であるヒドロキシベンゼンを構造的な上位概念としてひも付ける。また、オントロジー的な階層構造（化合物の上位下位に関する概念）とは別に、ヒドロキシベンゼンの官能基ごとの分類、添加剤や原料、製品などの用途、物性、法規制などの観点からも関連付けることで、複雑なグラフネットワークが構築されていく。

このように、単純に有機化合物をグラフ構造で表



SMILES : Simplified Molecular Input Line Entry System *
 ※分子の化学構造を英数字で文字列化した、構造の曖昧性が少ない表記方法

図-4 化合物ナレッジグラフの部分構造の例

すだけでなく、関連する化合物の特徴量、現象を含めてグラフ構造化することで、少ない物性情報からでも材料設計や解析に必要な特徴量を推定できるようになる。また、有機化合物で数多く存在する異名表記に対して、データベース間でオントロジーを考慮して同一性判定をすることによって、特許や論文などから化学物質を漏れなく探索できる。更に、反応のメカニズムの関係性を化学ナレッジグラフ上で一定の規則に従って学習することで、蓄積したナレッジに裏付けられた材料の効率的な探索、自動設計を目指していく。

6. むすび

本稿では、トラストで説明可能なAIのベースとなっているナレッジグラフの概要と、金融、化学の各分野におけるナレッジグラフ構築の取り組みについて紹介した。

説明可能なAIの研究開発が世界規模で行われる中、富士通研究所では、ナレッジグラフで表現された知識を基にして、AIによる推定結果に理由や根拠を付与することで、AIの信頼性を向上させるトラストで説明可能なAIを実現し、その社会実装を進めている。今後、富士通研究所は実際の適用分野を見据えた研究開発を通じて技術を磨き、様々な分野で期待される信頼性を持った「トラストで利用可能なAI」を提供していく。

本稿に掲載されている会社名・製品名は、各社所有の商標もしくは登録商標を含みます。

参考文献

- (1) 岡本青史ほか：思考を綴るAIがゲノム医療の新たなステージを拓く。BIO Clinica 34巻3号, p.48-54, 2019年3月。
- (2) 富士 秀ほか：Deep Tensorとナレッジグラフを融合した説明可能なAI。FUJITSU, Vol.69, No.4, p.90-96 (2018).
<https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol69-4/paper14.pdf>
- (3) 谷 聡史：トランザクションデータを活用したAIスコアリングモデル運用のライフサイクル。FUJITSU,

Vol.70, No.3, p.41-46 (2019).

<https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol70-3/paper07.pdf>

- (4) 国税庁, 法人番号公表サイト。
<https://www.houjin-bangou.nta.go.jp/>
- (5) 金融庁, 金融商品取引法に基づく有価証券報告書等の開示書類に関する電子開示システム。
<http://disclosure.edinet-fsa.go.jp/>
- (6) 田中一成ほか：化学データ利活用サービスに向けたナレッジグラフ構築。Japio YEAR BOOK 2018.
- (7) J-GLOBAL化学技術総合リンクセンター：日本化学物質辞書Web。
<https://jglobal.jst.go.jp/info/nikkaji>
- (8) NCBI PubChem BioAssay.
<https://www.ncbi.nlm.nih.gov/pcassay>

著者紹介



富士 秀 (ふじ まさる)

(株) 富士通研究所
人工知能研究所
ナレッジグラフ, 自然言語処理などの人工知能関連の研究に従事。



中澤 克仁 (なかざわ かつひと)

(株) 富士通研究所
人工知能研究所
金融ナレッジグラフの構築・活用, AIの品質管理技術の研究開発に従事。



吉田 宏章 (よしだ ひろあき)

(株) 富士通研究所
人工知能研究所
化学ナレッジグラフ, 機械学習の研究開発に従事。