

第3世代ニューラル翻訳技術と翻訳バンクの展開報告

NICT

隅田 英一郎

<https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>

みんなの自動翻訳@TexTra

 **みんなの自動翻訳@TexTra®**

「みんなの自動翻訳@TexTra®」は、自動翻訳をみんなで育てるサイトです。

Language ▼

☐ ログインしたままにする
[パスワード再発行](#)

このサイトについて



みんなの自動翻訳@TexTra®とは



「みんなの自動翻訳@TexTra®」は、国立研究開発法人情報通信研究機構（NICT）が開発した自動翻訳サイトです。

最新の自動翻訳研究に基づく「高精度自動翻訳エンジン」が無料でご利用いただけます。

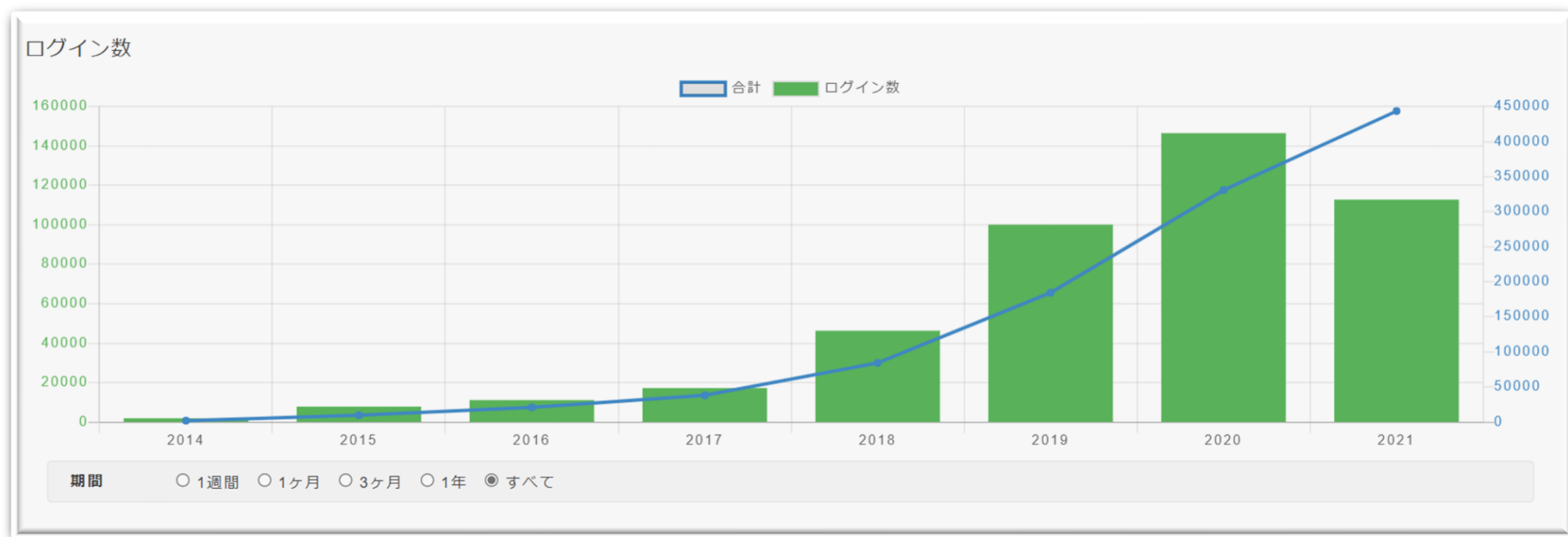
※利用登録に際しては、氏名・メールアドレスなどの個人情報は必要ありません。

便利な「翻訳エディタ」の利用も面倒なインストールは不要！

ウェブサイトやワードファイルなどを見ただけのままに読み込んで、自動翻訳し修正作業を行うことができる「翻訳エディタ」も簡単にご利用いただけます。

情報通信研究機構で公開している自動翻訳サイト
※本サイトのご利用は非商用利用に限定されます。
※メンテナンスのため定期的に止まります

多くの方にご利用いただいています



2021年11月9日

みんなの自動翻訳@TexTraの歴史

	ニュース@ https://mt-auto-minhon-mlt.ucri.jgn-x.jp/
2014年	一般公開開始(日英・日中・日韓)
2015年	科学技術文献データベースの作成に「高精度自動翻訳システム」を導入
2016年	NICTと特許庁の特許文献の機械翻訳に関する協力の継続について
2017年	ニューラル機械翻訳で音声翻訳アプリVoiceTraが更なる高精度化を実現
2018年	業界初、人工知能を活用した医薬業界向け自動翻訳システム、アストラゼネカと情報通信研究機構（NICT）で共同開発を実施
2019年	自動車法規文の自動翻訳をニューラル技術で高精度化
2020年	第2回日本オープンイノベーション大賞総務大臣賞受賞
2020年	オープンソースのコミュニティにNICT「みんなの自動翻訳」を提供

自動翻訳の歴史

- 計算機が発明されてその応用として自動翻訳の研究が開始
 - 第2次大戦後すぐにWarren Weaverが提案
- 3/4世紀の研究開発を経て、自動翻訳が一般に普及
 - ポケットーク等の音声翻訳専用機
 - VoiceTra等のスマホアプリ
 - GOOGLE翻訳
 - DeepL翻訳
 - みんなの自動翻訳@TexTra

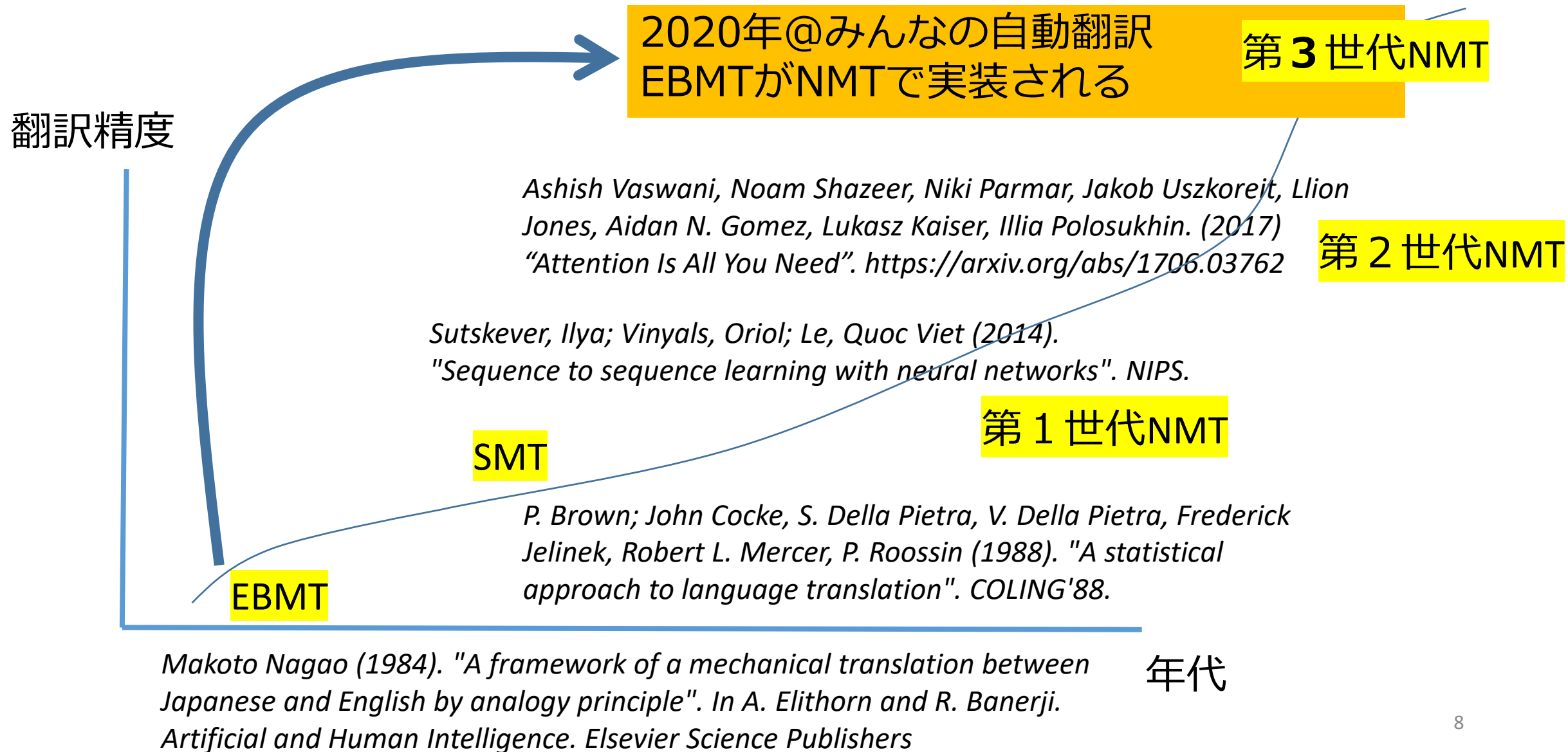
自動翻訳技術のタイプ

- ルールベース自動翻訳
文法規則や辞書を人間が記述
上記に基づき自動翻訳を実施
- コーパスベース自動翻訳（MT）
対訳コーパスから自動翻訳エンジンを自動学習
任意言語対に対して適用可能
ニューラル機械翻訳（NMT）もその一つ

対訳コーパスの一例

- Italy have defeated Portugal 31-5 in Pool C of the 2007 Rugby World Cup at Parc des Princes, Paris, France.
- Italia berhasil mengalahkan Portugal 31-5 di grup C dalam Piala Dunia Rugby 2007 di Parc des Princes, Paris, Perancis.
- フランスのパリ、パルク・デ・フランスで行われた2007年ラグビーワールドカップのプールCで、イタリアは31対5でポルトガルを下した。
- អ៊ីតាលីបានឈ្នះលើព័រទុយហ្គាល់ 31-5 ក្នុងប្លង់ C នៃពិធីប្រកួតពាន់រង្វាន់ពិភពលោកនៃកីឡាបាល់ឱបឆ្នាំ 2007 ដែលប្រព្រឹត្តទៅប៉ារីស ខេត្តសេន ប្រីន ក្រុងប៉ារីស បារាំង។
- Itali telah mengalahkan Portugal 31-5 dalam Pool C pada Piala Dunia Ragbi 2007 di Parc des Princes, Paris, Perancis.
- ပြင်သစ်နိုင်ငံ ပါရီမြို့ ပါ့ဒက်စ် ပူရင့်စက် ၌ ၂၀၀၇ခုနှစ် ရပ်ဘီ ကမ္ဘာ့ ဖလား တွင် အီတလီ သည် ပေါ်တူဂီ ကို ၃၁-၅ ဂိုး ဖြင့် ရေကူးကန် စီ တွင် ရှုံးနိမ့်သွားပါသည်။ ။
- Ý đã đánh bại Bồ Đào Nha với tỉ số 31-5 ở Bảng C Giải vô địch Rugby thế giới 2007 tại Parc des Princes, Pari, Pháp.
- อิตาลีได้เอาชนะโปรตุเกสด้วยคะแนน31ต่อ5 ในกลุ่มc ของการแข่งขันรักบี้เวิลด์คัพปี2007 ที่สนามปาร์กเดแพร็งส์ ที่กรุงปารีส ประเทศฝรั่งเศส
- Natalo ng Italya ang Portugal sa puntos na 31-5 sa Grupong C noong 2007 sa Pandaigdigang laro ng Ragbi sa Parc des Princes, Paris, France.

コーパスベースMTアルゴリズムの進展



第1世代と第2世代の自動評価尺度BLEUの比較

	汎用日英	汎用英日	特許日英	特許英日	
2017	22.1	24.2	40.7	41.7	第1世代
2018	21.2	24.5	40.0	41.6	第1世代
2019	20.8	23.2	37.7	40.1	第1世代
2020	27.6	31.5	45.6	46.8	第2世代
2021	28.1	31.0	47.3	46.5	第2世代

汎用・特許ともに、

第1世代から第2世代にかけて、大きくBLEU値が向上

(BLEUは性能の目安で30程度以上なら一般的に使える翻訳)

訓練・アダプテーション・EBMT

【訓練】 1文ずつNMTモデルのパラメタを調整する

• 大雑把には：

- 入力文を翻訳
- 参照訳文と比較
- 翻訳文と参照訳文の違いに応じてNMTのパラメタを更新
- 以上を大規模に繰り返す（数億回になることもある）

【アダプテーション】（fine tuning とも言います）

- 訓練済みNMTモデルに、上記訓練を特定分野データで追加
- 訓練済みモデルをベースにするので、比較的少量データで高精度

【EBMT】（NICT開発・詳細未発表）

- 入力文と類似した対訳文をデータベースから検索
- 十分に類似した文があるときには、それを参考に自動翻訳
- 類似文がない場合には、ベースのNMTで自動翻訳

金融分野でのアダプテーション・EBMT

	汎用	ADAPT	ADAPT+EBMT
訓練と重複あり1000	31.0	38.1	55.2
訓練と重複無し514	29.3	35.2	49.1
訓練と重複のみ416文	34.4	44.1	67.5

適時開示文書は、異なる文書であっても重複する文が多いので、EBMTの効果が著しく高い

オープンソースのコミュニティにNICT 「みんなの自動翻訳」を提供

2020年12月2日

国立研究開発法人情報通信研究機構
Linux Foundation Japan

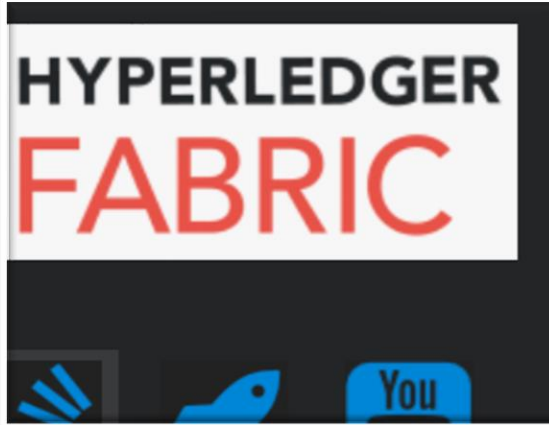
ツイート

いいね！ 25

国立研究開発法人情報通信研究機構（理事長：徳田 英幸、以下「NICT」）は、Linuxを含むオープンソースソフトウェアを推進するLinux Foundation Japan（日本代表者：福安 徳晃、以下「LF」）に、「みんなの自動翻訳」のAI自動翻訳技術を提供します。LFは、AI自動翻訳技術をLFのサーバに実装し、2021年1～3月期のサービス開始を目指します。LFのオープンソースプロジェクトで同自動翻訳エンジンを使用し、翻訳作業で生成されるオープンソース技術関連の対訳・語彙データをNICTの[翻訳バンク](#)^{注1}にフィードバックします。

両者が協力することで、オープンソースソフトウェア向けの高精度AI自動翻訳システム^{注2}を育てていくエコシステムを創出し、日本のソフトウェア開発を支援してまいります。

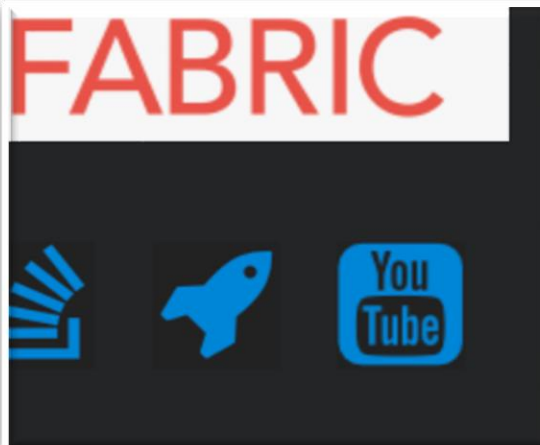
翻訳エディタの修正からエンジンを学習



Introduction

In general terms, a blockchain is an immutable transaction ledger, maintained within a distributed network of *peer nodes*. These nodes each maintain a copy of the ledger by applying transactions that have been validated by a *consensus protocol*, grouped into blocks that include a hash that bind each block to the preceding block.

<https://hyperledger-fabric.readthedocs.io/en/release-2.2/whatis.html>



一般的に、ブロックチェーンはイミュータブルなトランザクションの台帳であり、複数のピアノードからなる分散ネットワークで管理されるものです。トランザクションは、ある合意形成プロトコルに従って検証され、ブロックにまとめられます。各ノードは、それぞれが持つ台帳のコピーに対して、これらのトランザクションを適用していくことで管理しています。なお、各ブロックは、直前のブロックと紐づけるためにハッシュ値を含んでいます。

<https://hyperledger-fabric.readthedocs.io/ja/release-2.2/whatis.html>

汎用NT+エンジンに日次で学習を反映

英語 → 日本語 ⇄

汎用NT+(実験... 1

翻訳

訳文コピー

←

オプション

ヘルプ

In general terms, a blockchain is an immutable transaction ledger, maintained within a distributed network of peer nodes.

These nodes each maintain a copy of the ledger by applying transactions that have been validated by a consensus protocol, grouped into blocks that include a hash that bind each block to the preceding block.

翻訳不要記号を挿入

一般的に、ブロックチェーンはイミュータブルなトランザクションの台帳であり、複数のピアノードからなる分散ネットワークで管理されるものです。

トランザクションは、ある合意形成プロトコルに従って検証され、ブロックにまとめられます。各ノードは、それぞれが持つ台帳のコピーに対して、これらのトランザクションを適用していくことで管理しています。なお、各ブロックは、直前のブロックと紐づけるためにハッシュ値を含ん

下記ページの冒頭を自動翻訳（2021年11月9日）

<https://hyperledger-fabric.readthedocs.io/en/release-2.2/whatis.html>

OSS翻訳に適した利用規約

1 アプリケーション利用者は、商用目的や公序良俗に反する目的のためにアプリケーションを利用することはできません。ただし、アプリケーションを利用して、Wikipedia等のCreative Commons License の文書や、GNU General Public License, MIT License, Apache License, BSD License 等のオープンソースソフトウェアの文書等の公益性が高い文書であって、アプリケーション利用者がその文書の自動翻訳結果や修正結果を利用者や目的の制限なく一般に無償公開する場合には、当該文書のライセンスが当該文書の利用の目的や範囲等を限定していない場合であっても、これによらず利用できます。

<https://mt-auto-minhon-mlt.ucri.jgn-x.jp/content/policy/>

OSS翻訳でのみんなの自動翻訳の利用例

オープンソースカンファレンス2021 Online/Spring セミナープログラム

OmegaTとみんなの自動翻訳でLibreOfficeの翻訳を爆速にした話

2021年3月6日(土) 11:00 ~ 11:45

セッション概要

国立研究開発法人情報通信研究機構(NICT)が開発している「みんなの自動翻訳」の学習データをOmegaT上に展開することで、長大な(英語版で545ページあります) LibreOffice 7.0 Calc Guide日本語版ドラフトを驚異的な速度で完成することができました。この成果を水平展開し、翻訳の時間・人員などが不足しているコミュニティにも活用してもらえよう、OmegaTの設定、学習データの導入およびフィードバックについて解説を行います。

講師

のがたじゅん

文化への展開

日越茶道・文化交流協会（所在地：東京都千代田区、会長：千玄室）が発足し翻訳バンク活動の文化面への支援を開始した。



<https://sites.google.com/view/vjca2023/>



日越茶道・文化交流協会 会長と特別顧問のご紹介



会長
千玄室

茶道界千宗第15代家元

1923年、京都府生まれ。1964年10月、第千宗第15代家元となり今日日庵庵主として宗室を襲名。2002年12月、瑞興千宗に家元を譲渡し、現職千玄室大僧正となる（梵号博士、文学博士）。

外務省参事、ユネスコ親善大使、日本・国連親善大使（外務省）。



特別顧問
寺島実郎

一般財団法人日本総合研究所 会長、多摩大学 学長

1947年生まれ。日本の評論家。多摩大学学長・教授、東北公益文科大学客員教授、立命館大学国際関係学部客員教授、帝京山大学特別客員教授、立命館アジア太平洋大学国際経営学部客員教授、学校法人市川学園特別顧問、一般財団法人日本総合研究所会長、一般社団法人寺島文庫代表理事。