

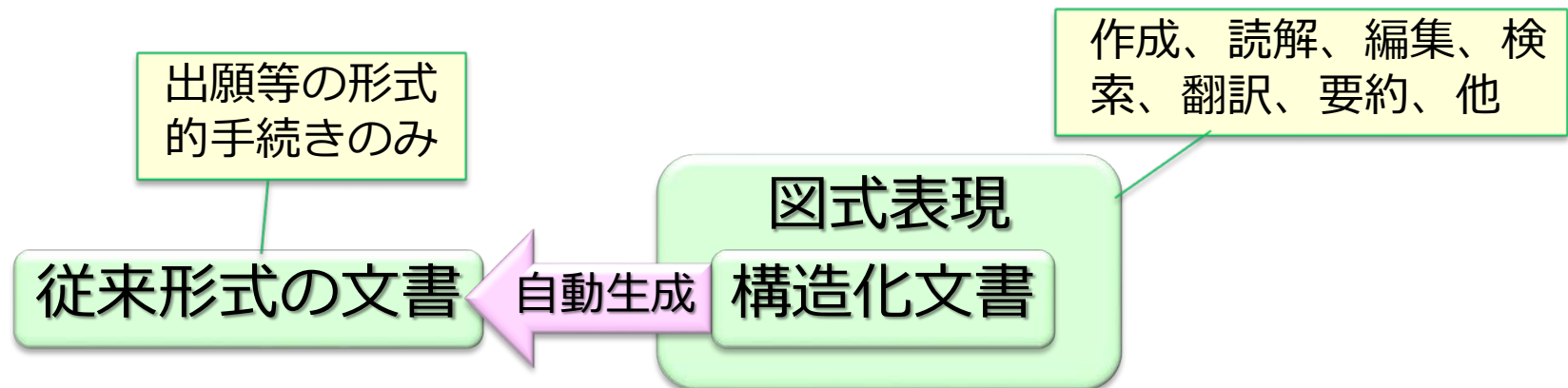
# 文書作成支援分科会報告

# AIの基盤としてのデータ整備


- AIによる社会の自動化にはデータ整備が必須
  - ◆ 研究開発と実用の両方で、良質のデータが容易に取得できる必要あり
    - \* 研究開発コストのほとんどがデータ整備にかかる
    - \* サービスの受容者に関する詳しいデータが必要
- データ整備 = BPR (業務改革)
  - ◆ 意味構造化データが潤沢に流通する社会の構築
    - \* 意味構造化 = 機械にも人間にも意味がわかる
- 業務改革には意味の理解が必要だが、意味を理解しないAIによるデータ整備の自動化は不可能

# 構造化文書を原本とする文書処理

- 構造化文書
  - ◆ 構造共有により表現を簡単に
    - \* 共参照はなるべく構造共有(図式表現ではリンク)で表わす
  - ◆ 従来の文書の意味を完全に表現できる
- 任意の文書について、意味が等しい文書を構造化文書から自動生成できる。
- 従来形式の文書は出願等の形式的手続きのみに用いる。



# 構造化文書（特許版）

- 特許は、主に【請求の範囲】と【明細書】、【図面】からなり、各文書は互いに関連し、様々な構造を持つ(発明の構造、請求項の引用関係、請求項と対応する実施の形態等)
  - 特許書類の記述形式
    - ◆ 素のテキスト(特許公開公報等)
      - \* 文書の構造が明確ではない=情報の2次利用が困難
    - ◆ タグ付けしたテキスト(XML)
      - \* 文書の構造を明確に記述可能
      - \* 冗長な記述が多く直接記述することが難しい
- 
- 特許作成支援に用いることができる新しい記述形式を提案
    - ◆ 特許文書の「ソースコード」
      - \* 特許の構造を明確かつ簡単に表現
      - \* 出願形式や図式表現等の他の表現への変換が可能
      - \* 訓練を受けた人間ならば直接記述することが容易に可能(TeXのソースファイルのようなもの)

# 構造化文書（特許版）

# 請求項 1

**claim** c.鉛筆 (鉛筆, 消しゴム, 連結具) **embodiments** e.鉛筆  
鉛筆と、(鉛筆芯, 鉛筆軸)  
鉛筆芯と鉛筆軸を有する  
消しゴムと、  
{鉛筆芯}で筆記された1以上の線を消去する  
連結具と、  
{鉛筆}の端と{消しゴム}の端とを連結する  
を備えた筆記具。

**effect**

鉛筆と消しゴムが一体となり、鉛筆の向きを反転させるだけで  
書き損じ部分に消しゴムを当てることが可能となる。

⋮

# 実施の形態

**embodiment** e.鉛筆 **figures** f.鉛筆  
**title** (実施の形態{num: \_\_self\_\_})  
' 本実施の形態において、{inst1}について説明する。  
' {inst1}は、{pen1}、{eraser1}、{connector1}を備える。

**component** pen1  
筆記具・文房具の一種であり、

⋮

赤太字	キーワード
緑斜体	変数定義
橙	変数参照
青	自動生成行
灰	コメント

# 図式表現

筆記具・文房具

顔料を細長く固めた芯（鉛筆芯）を軸（鉛筆軸）ではさんで持ち易くしたもの

通常、黒鉛と粘土を練り合わせて焼き固めたもの

叙述

定義

叙述

鉛筆芯

構成要素

鉛筆

構成要素

鉛筆軸

叙述

通常、木材であるが、古紙などでもよい

請求項

筆記具

構成要素

構成要素

消しゴム

構成要素

連結具

機能

○の端と○の端を連結

機能

主に鉛筆などで書かれたものを消去

実施の形態

叙述

鉛筆芯の方向と垂直な面に対する凸包（外周）の一部が直線となる形状とすることは好適

例

外周を六角形とすることで、筆記具を机に置いた場合に転がり防止できる

叙述

文房具

原料

研磨材

原料

サブスティチュート

原料

天然の生ゴム

# 構造化用語データUTX

- 文書内の用語の定義と階層を規定し管理できる。
- 詳細は <http://www.aamt.info/Japanese/utx/>

#UTX 1.20; lang: ja/en-US; last modified date: 2017-01-28										
#日本語	英語	品詞	上位語	下位語	上位構成要素	下位構成要素	概念ID	定義	叙述	コメント
term:ja	term:en	pos:ja	hypernym:ja	hyponym:ja	partOf:ja	components:ja	concept ID	x-definition		
文房具	stationery	noun		鉛筆			1			
鉛筆	pencil	noun	文房具			鉛筆芯, 鉛筆軸	2	筆記具・文房具の一種であり、顔料を細長く固めた芯(鉛筆芯)を軸(鉛筆軸)ではさんで持ち易くしたものの		
鉛筆芯	lead	noun			鉛筆		3	顔料を細長く固めた芯。	通常、黒鉛と粘土を練り合わせて焼き固めたもの	
鉛筆軸	body	noun			鉛筆		4		通常、木材であるが、古紙などでもよい	
シャーペン	Sharpen	noun				シャープペン, 消しゴム, 連結具	5			
シャープペン	sharp pen	noun			シャーペン		6	黒鉛の芯を随時に補充可能な機械式の筆記具		
消しゴム	eraser	noun			シャーペン		7	主に鉛筆などで書かれたものを消去するときに使う文房具		
連結具	connector	noun			シャーペン		8			
連結する	connect	verb					9	ひと続きになるように繋ぎ合わせる		
線	line	noun				線, 塗りつぶし, 文字, 絵				

# 翻訳支援ツールによる用語検証

作成した用語データを用いて訳文の用語をチェックできる。検索にも活用可。

The screenshot shows the SDL Trados Studio interface with the Terminology Verifier window open. The window displays a list of 138 messages, with 15 errors and 123 warnings. The messages are categorized by severity and include details about term usage and expected translations. A red box highlights the first three error messages, and a red arrow points from the Terminology Verifier window to the expanded view below.

Severity	Message	Origin
✖	Wrong usage of the term "エントリー" - this term is defined as forbidden.	Terminology Verifier
✖	Wrong usage of the term "エントリー" - this term is defined as forbidden.	Terminology Verifier
✖	Wrong usage of the term "エントリー" - this term is defined as forbidden.	Terminology Verifier
⚠	Target segment does not contain the translated term for "user role" - "ユーザー ロール" expected.	Terminology Verifier
⚠	Target segment does not contain the translated term for "user role" - "ユーザー ロール" expected.	Terminology Verifier
⚠	Target segment does not contain the translated term for "unidirectional" - "一方向" expected.	Terminology Verifier
⚠	Target segment does not contain the translated term for "unidirectional" - "一方向" expected.	Terminology Verifier
⚠	Target segment does not contain the translated term for "term" - "用語" expected.	Terminology Verifier
⚠	Target segment does not contain the translated term for "term" - "用語" expected.	Terminology Verifier

✖	Wrong usage of the term "エントリー" - this term is defined as forbidden.
✖	Wrong usage of the term "エントリー" - this term is defined as forbidden.
✖	Wrong usage of the term "エントリー" - this term is defined as forbidden.
⚠	Target segment does not contain the translated term for "user role" - "ユーザー ロール" expected.
⚠	Target segment does not contain the translated term for "user role" - "ユーザー ロール" expected.
⚠	Target segment does not contain the translated term for "unidirectional" - "一方向" expected.
⚠	Target segment does not contain the translated term for "unidirectional" - "一方向" expected.
⚠	Target segment does not contain the translated term for "term" - "用語" expected.
⚠	Target segment does not contain the translated term for "term" - "用語" expected.