

保険関連文書を対象とした 校正支援システム

(株)ミック インターンシップ
長岡技術科学大学 電気系
大平真一

はじめに

- 保険分野でも組版ソフトなどの利用が増加
- 校正は従来通り人手で行われている
- 誤りが許されないため校正の**コストが大きい**
- 精度を維持しつつコストを下げる必要がある

保険関連文書とは

- 基礎書類
 - 契約内容の詳細を記載した書類
 - 普通約款
 - 特別約款
- 派生書類
 - 消費者への説明に用いられる書類
 - パンフレット
 - 説明書

基礎書類とは

- 契約条項が記載されている
- 省庁に認可を受ける必要がある
- 特徴
 - 法律分に近い文体
 - 章・条・項などの構造が厳密
 - 図や表は補助的に使用される

派生書類とは

- 説明書やパンフレット等様々な形態
- わかりやすさを重視して作成される
- 特徴
 - 簡単な語や表現の使用
 - 図や表を中心とした構成

ガイドライン

- 送り仮名や常用外漢字の扱い
- 読みにくさの要因
 - 複雑な文章構造
 - 二重否定や入り組んだ括弧書きなど
 - 難解な用語
 - 保険や医療に関する専門用語など
 - 1文中での同音単語の使用

保険会社

消費者

金融庁

基礎書類

標準約款

特別約款

事業方法書

派生書類

重要事項
説明書

規定集

パンフレット

etc.

申請

認可

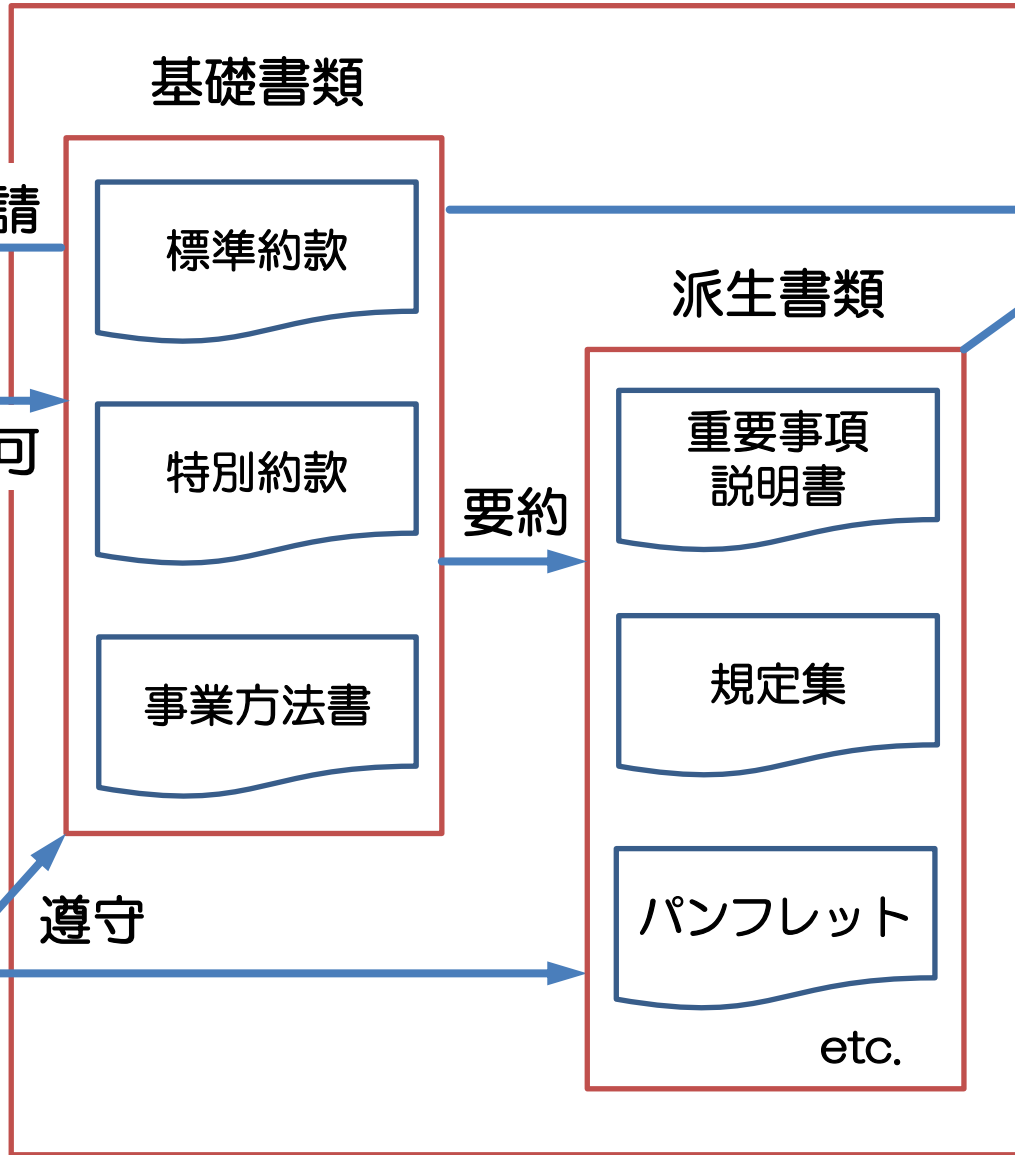
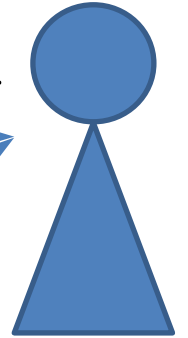
要約

配布

遵守

各種保険協会

ガイドライン



派生書類の校正

- 内容についての矛盾の解消
 - 基礎書類内の対応する文の参照
- 語の誤り訂正
 - 変換誤りやガイドラインの遵守の確認

内容についての矛盾の解消

- 基礎書類の対応部分の参照
 - 量的な問題
 - 基礎書類は数百ページにわたる
 - 派生書類との複雑な対応関係
 - 1文と基礎書類の複数文の対応が存在
 - 構造の違いから広範囲の確認が必要

語の誤り訂正

- 人手で検出されにくい誤り
 - 表記ゆれ
 - 送り仮名
 - かなと漢字の混在(例.「補綴」→「補てつ」)
 - 誤字
 - 同音異字語の変換誤り(例.「障害」・「傷害」)
 - 専門用語の誤り

システムの概要

- 基礎書類と派生書類の自動対応付け
 - 類似度計算を用いた自動対応付け
- 誤り検出
 - 類似文での語の使われ方から誤りを推定
 - ガイドライン遵守のチェック

自動対応付け

- 内容語の一致率を用いた類似度計算
 - 頻度による重み付け(低頻度語:重み大)
- 類似度の大きいものを対応する文として表示
- 内容語の数が大きく異なる場合を考慮
 - Simpson係数とJaccard係数をベースにした計算式

誤り検出

- 見逃されやすい誤り
 - 表記ゆれや同音異字などの変換ミス
- ガイドラインの遵守
- 類似文中での同音単語の使われ方を用いる例.

入力文「...業務上の災害を**保証**する...」

類似文「...災害を**補償**する他の災害**補償**制度...」



出力文「...業務上の災害を**補償**する...」

保険関連文書以外への適応について

- 基礎・派生書類に類似した関係の文書
 - 金融や医薬品に関する文書
- 独自のガイドラインを持つ文書
 - 社内向けの文書など