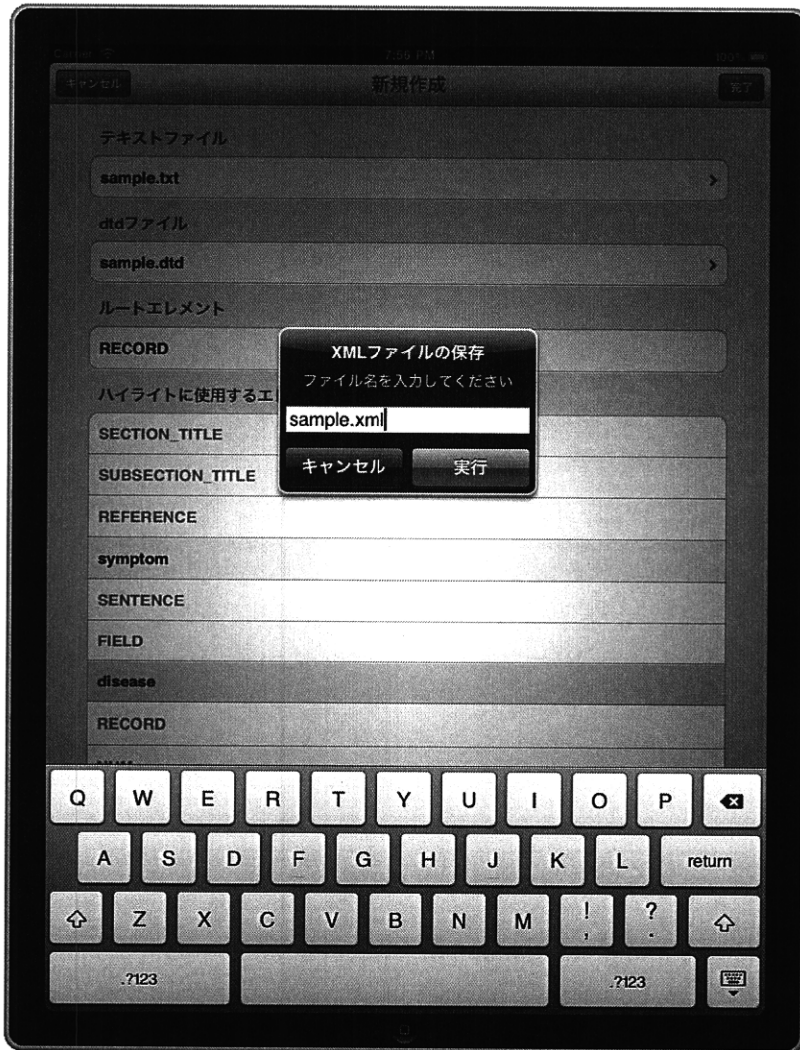


新規XMLファイルの生成



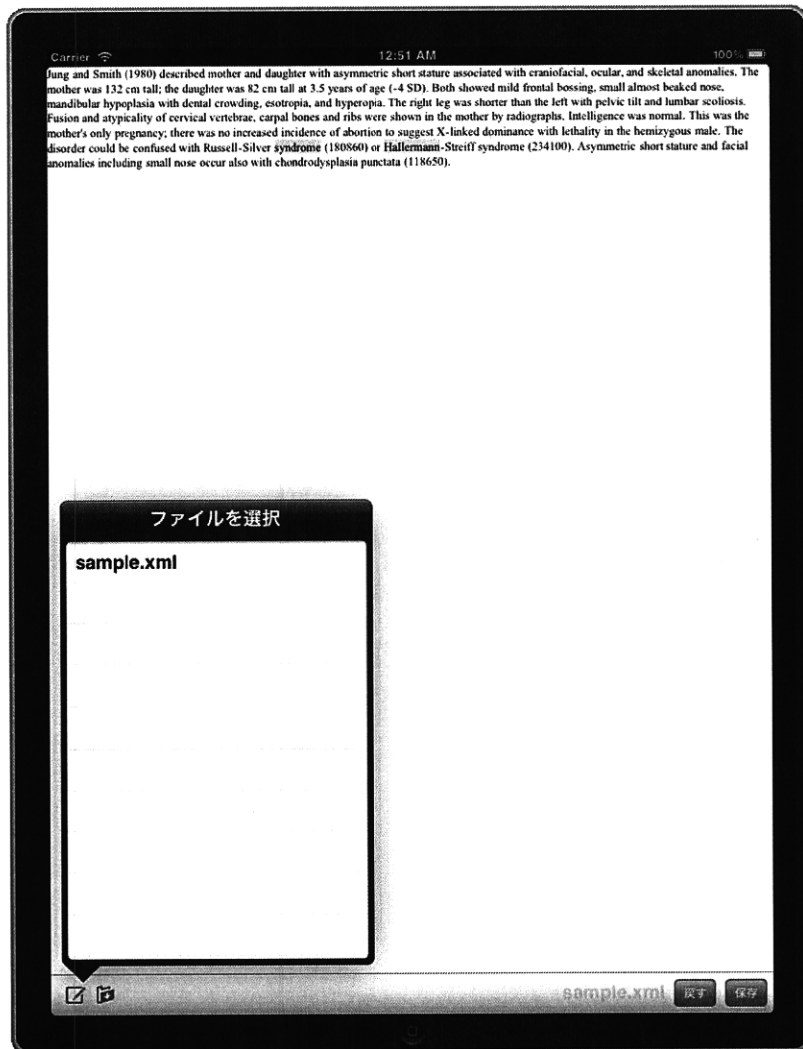
ひとつおりの設定がおわったら、完了ボタンをほして、XMLファイルを生成します。
作成するXMLファイル名を入力してください。
XMLファイルが生成されると、編集画面に移行します。

ファイルの上書き確認



既に同名のファイルがある場合は上書き確認を行いますので注意して作成してください。

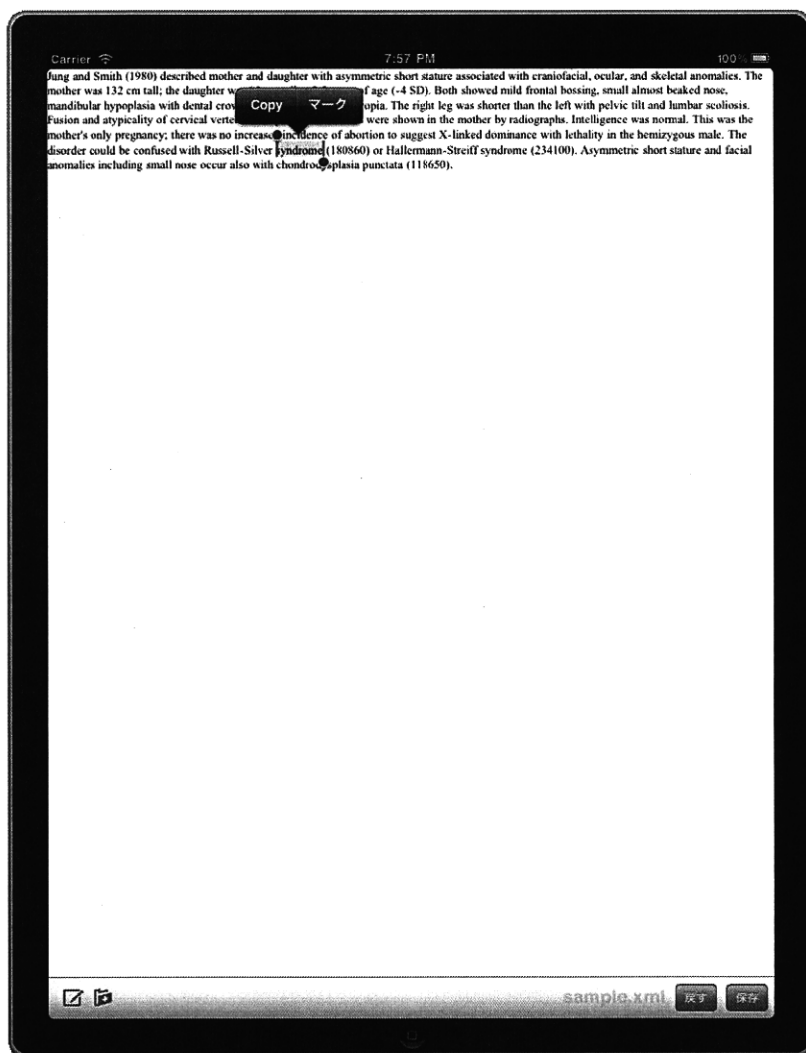
既存XMLファイルを開く



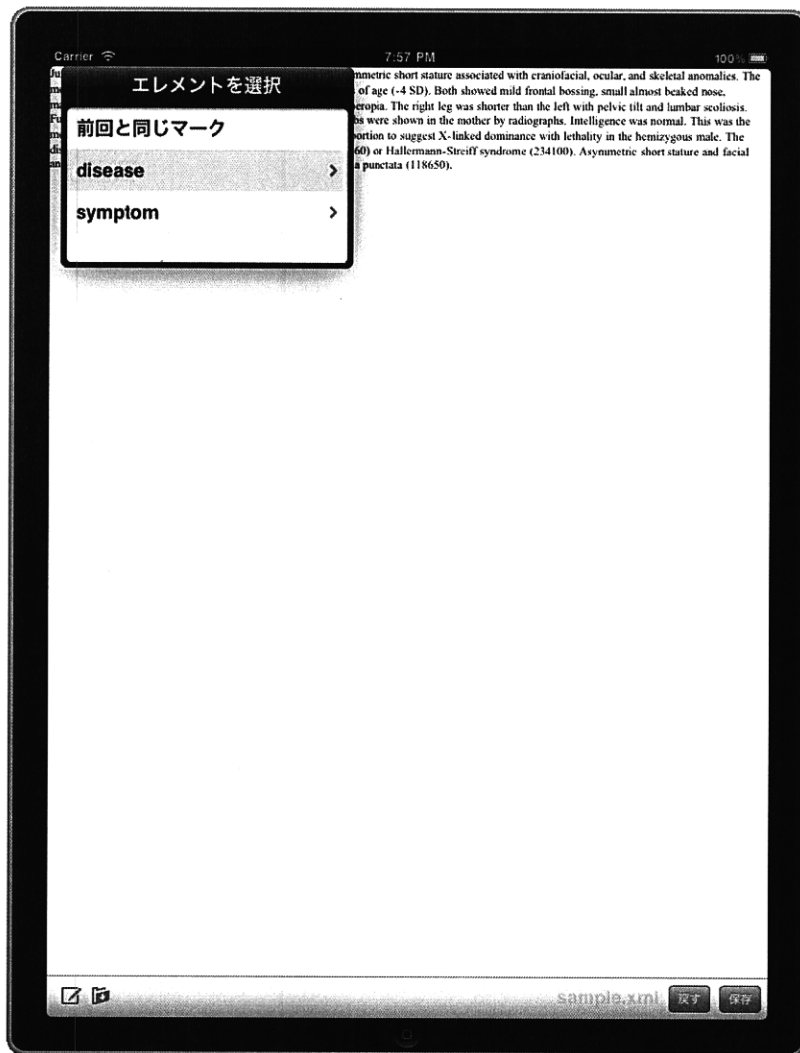
他のファイルを編集したい場合はファイルを開くアイコンをタッチしてファイルを選択してください。選択したファイルを開いて編集画面に移行します。

編集画面の操作

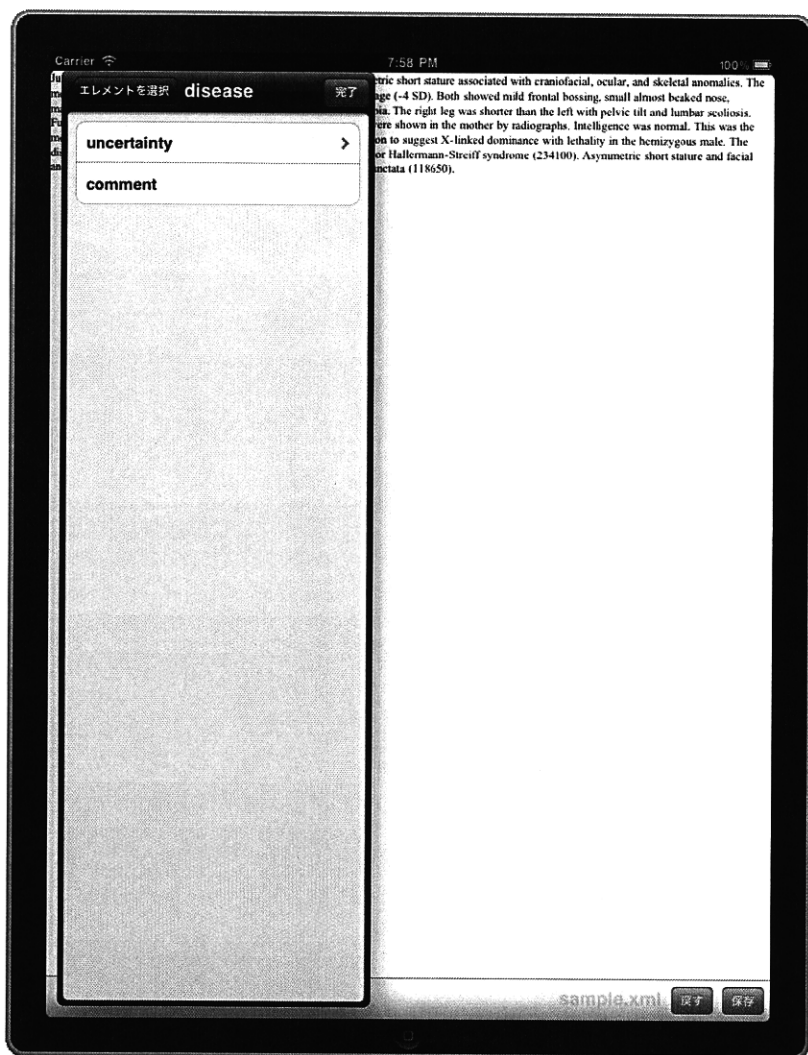
マーキング



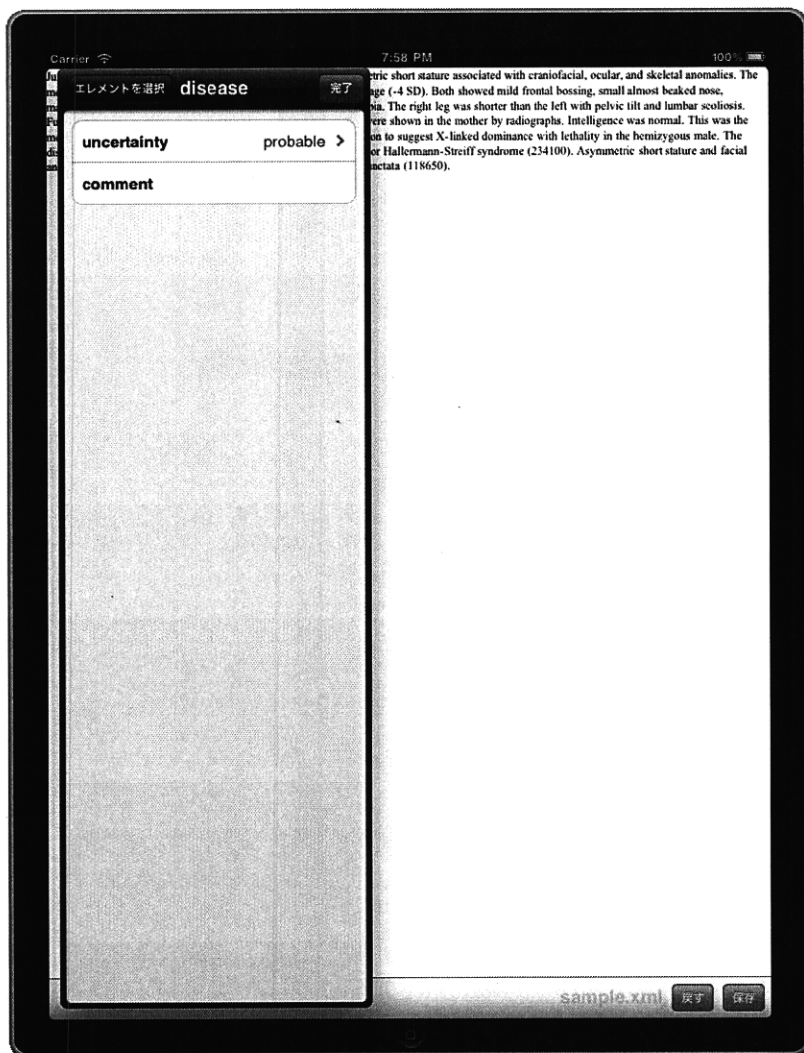
タグ付けしたい部分を選択します。
選択したい付近でタッチし長押しすると範囲選択モードに移行しますので、選択してください。



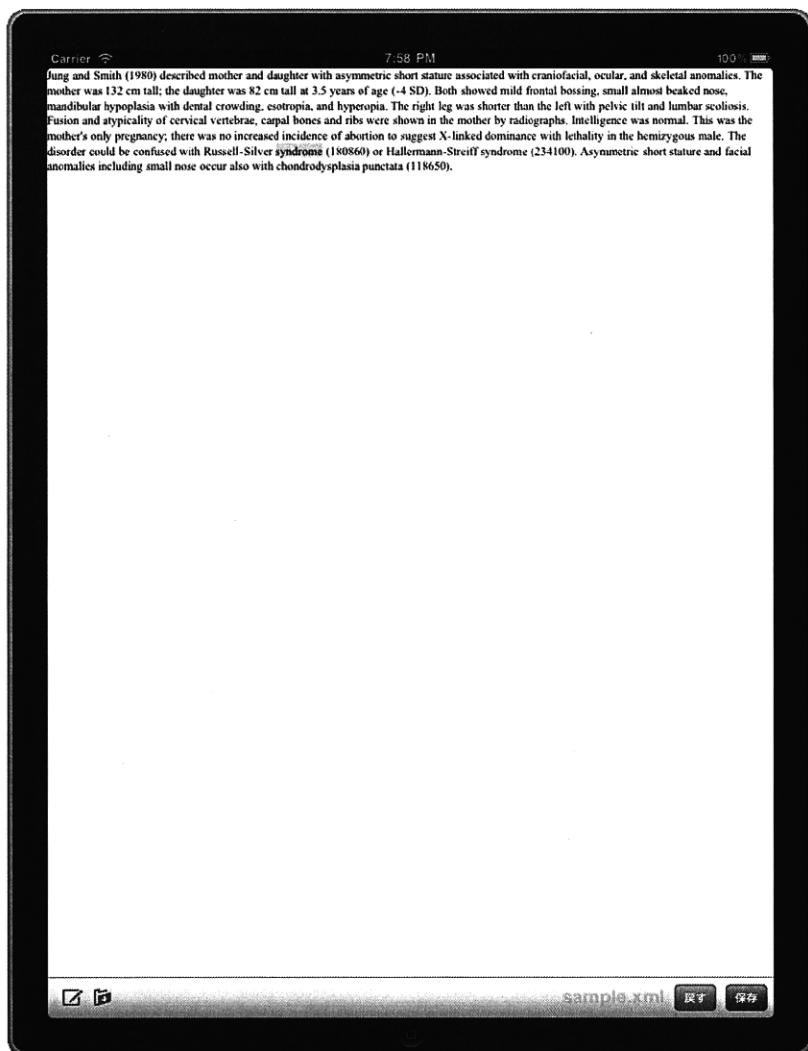
選択すると、タグの一覧が表示されます。適応するタグを選択してください。



値が選択可能な属性をもつエレメントの場合は属性の選択画面が表示されます。
(選択は任意です。)

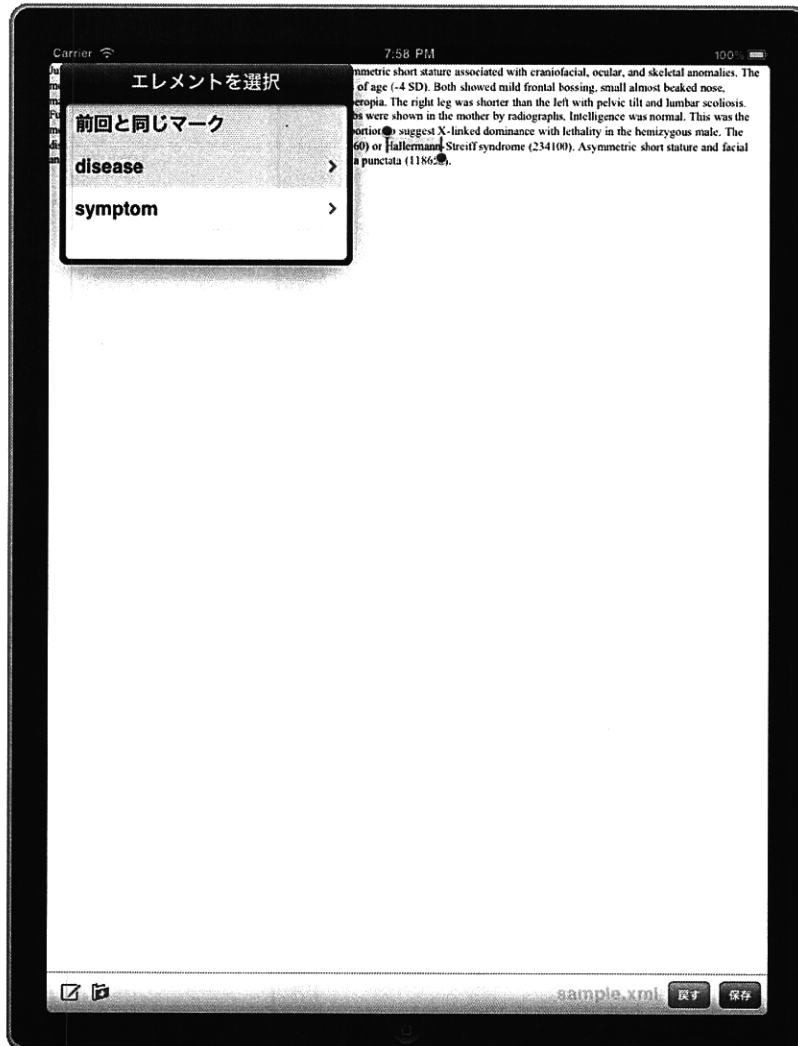


選択がおわったら完了ボタンをタッチしてください。



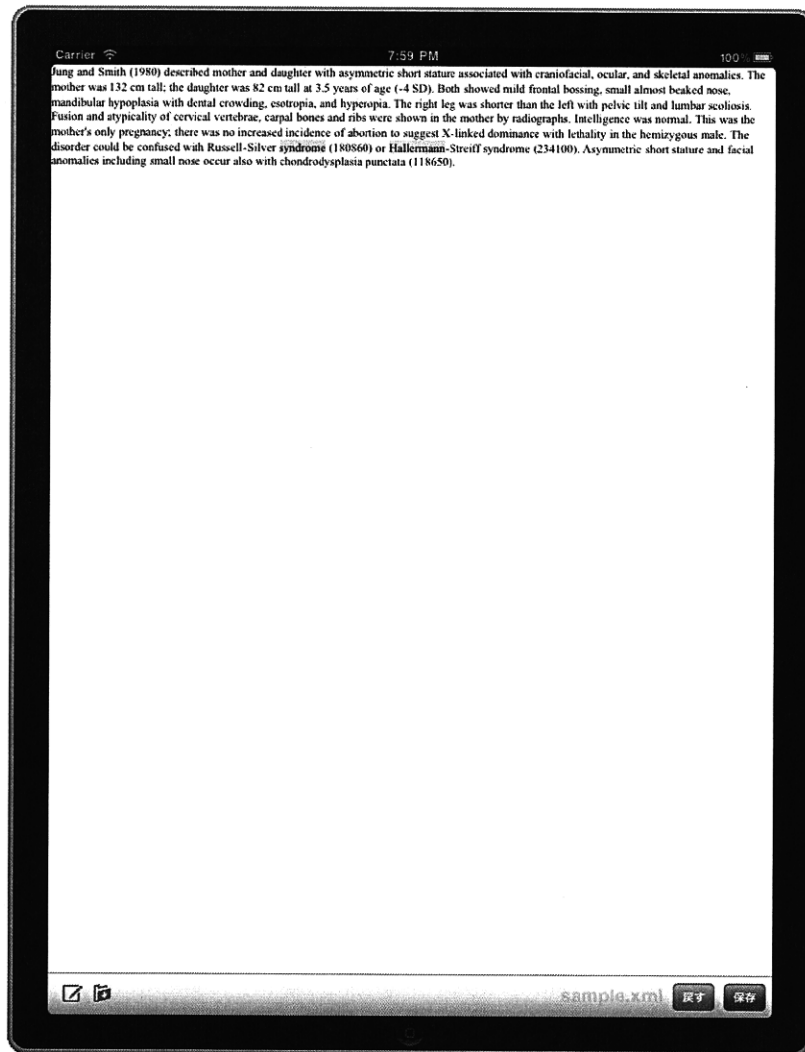
選択した範囲が指定の色でハイライトされます。

前回と同じマーク



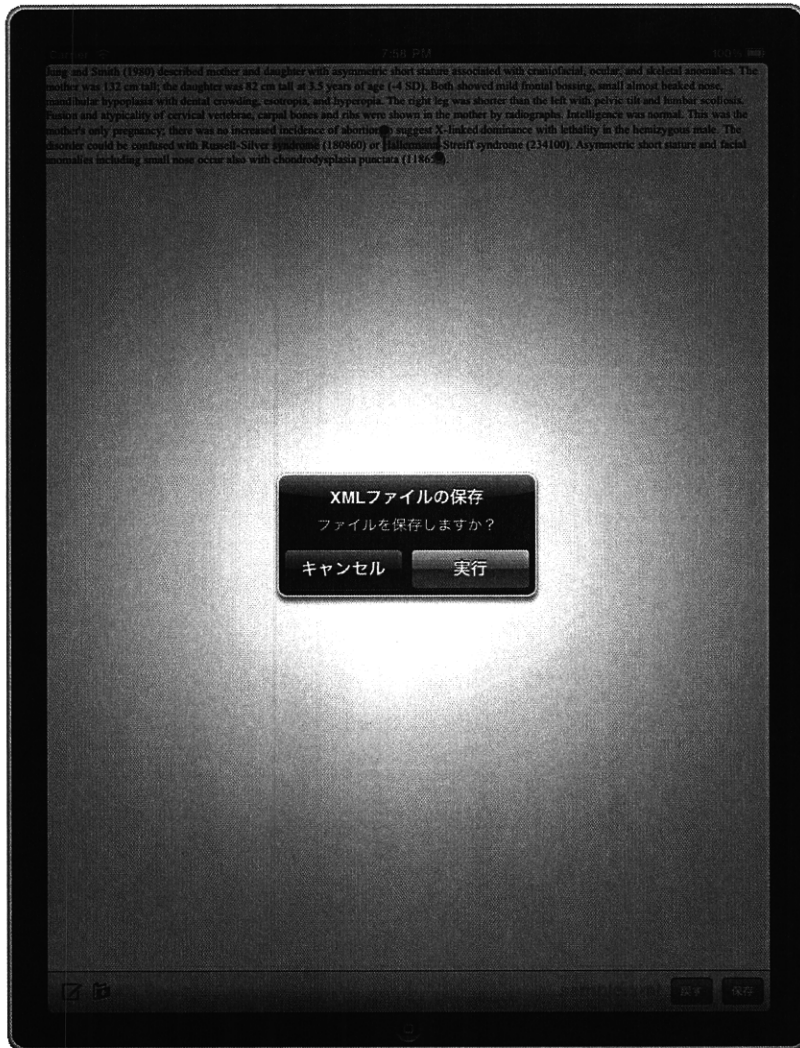
「前回と同じマーク」を選択すると直前に指定したのと同内容でタグ付けされます。同じタグ付けを繰り返す場合に利用してください。

取り消し



画面右下の「戻す」をタッチすると直前のタグ付けが取り消されます。取り消しは繰り返し実行可能です。（ファイルを保存した段階まで。）

保存



タグ付けした内容をファイルに反映させるには画面右下の「保存」をタッチしてファイルを保存してください。（「保存」を実行すると「取り消し」はできなくなりますので注意してください。）

III-7. 疾患知識ベースの構築戦略

【報告書】

疾患記載の指摘、分離、分類（E-1）

2011/03/15

アメリエフ株式会社

1. 概要

臨床試験の検査結果と疾患の関係を示したデータに対して、各検査結果から想定される疾患のリストを具体的な病名に展開することが目的である。第一段階として、疾患記載の指摘、分離、分離を行うための作業手順を提案する。

作業手順は、除外作業（処理が必要ないデータおよび文章）、前・中・後処理、目視確認およびデータの統合の流れで行う。作業内容は、不要部分の削除、疾患名への展開、補足項目に関する処理の3つに大別した。

2. 手順

別紙の要旨を以下にまとめた。数字は、E-1_excluded_110322_strategy.xlsx の「グループ化・テキストフィルタ」sheet を元に算出。

i. 処理が必要ないデータを除外

処理が必要ない行やセルを除く。

ii. 文章を除外

ひらがなを7文字以上含む記載を文章と考えて、取り出し別テーブルとする。（A 参照）
中処理の決定に従って、主に専門家による手動処理で修正を行う。

iii. 前処理

不要部分の削除を、手動処理および自動処理で行う。（B1,B2 参照）

疾患名への展開を、主に専門家による手動処理で行う。（C1,C2 参照）

iv. 中処理

補足事項に関する表記方法を決定する。（D1,D2,D3,D4 参照）

決定に従って、主に自動処理で修正を行う。

v. 後処理

前処理および中処理で処理できなかった、不要部分の削除、疾患名への展開、補足項目に関する処理を行う。

カッコを含む記載に対して、主に専門家による手動処理で上記の処理を行う。(E 参照)
ひらがなを含まない記載をこれ以上処理が必要ない記載と考えて、取り出し別テーブルとする。(F 参照)

残りの、ひらがなを 1 文字以上 7 文字以下含む記載から、一部の助詞を含む記載を取り出し、主に専門家による手動処理で行う。(G 参照)

ひらがなを 1 文字以上 7 文字以下含む記載を、取り出し別テーブルとする。(H 参照)

vi. 目視確認およびデータの統合

3つのテーブル(A,F,H 参照)に対して目視確認を行い、指定の形式に変換する。

3. まとめ

上記手順により修正される項目は 2093 カ所であり、1148 要素が修正される。

手動処理は 2 項目、自動処理は 20 項目である。

専門家による手動処理は 6 項目であり、作業時間としては約 173 時間(前処理 2160 分、後処理 5490 分、目視確認 2736 分)を要する見込みである。

