グラフ形式に基づく文書作成支援

2014-02-27

東京大学 大学院情報理工学系研究科
ソーシャルICT研究センター
教授　橋田 浩一

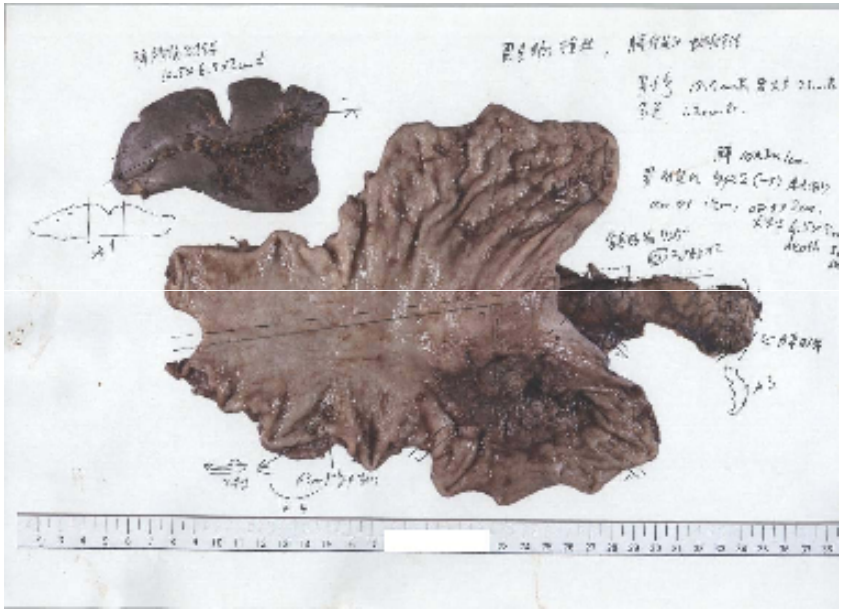
概要

- 目的
  - 文書データの二次利用
    - 検索、分析、翻訳、要約、他
  - 前提としての一次利用
    - 創作、記録、伝達、他
- 手段
  - グラフ形式により文書の意味構造を明示
  - 意味構造の明示により文書の一次利用(作成・読解)を支援
    - 少なくとも邪魔にはならない

文書データのグラフ形式 ⊃ 木形式

- 人間にもコンピュータにも扱いやすい
  - 人間による理解と作成が容易
  - 自動解析の精度が高い
- データの一次利用
  - 記録作業(文書作成)のコストを低減
  - 文書の読解を容易にし誤読を防止
- データの二次利用
  - 標準オントロジーによる正確な意味的構造化
  - 検索、要約、分析、翻訳などの自動化

病理診断報告書(従来形式)



胃全摘検体。小弯長15.7cm、大弯長23cm。口側周径3cm、肛門側周径4.5cm。1.2cm長の食道が付いている。

胃の上部後壁に6.5x5cm大の２型腫瘍がある。腫瘍表面は褐色調で、3x2cm大の潰瘍を有している。肛門側断端からは11cm、口側断端からは2cm離れている。腫瘍は各断端に及んでいない。腫瘍の漿膜側は硬くなっているが、腫瘍の露出は見られない。腫瘍は割面では白色充実性で、出血や壊死は僅かである。

以上の主病変とは別に、下部大弯に3mm大の亜有茎性ポリープがある。肛門側断端からは6cm離れている。

そのほかの粘膜面に著変はみられない。

人間にとってもコンピュータにとっても扱いにくい。

病理診断報告書(木形式)

凡例: 主題
主題への参照

最左の行は検体

検体の大弯

胃全摘検体

- 小弯長16cm、大弯長23cm
- 口側周径3cm、肛門側周径4.5cm
- 1.2cm長の食道が付いている
- 上部後壁に6.5x5cm大の2型腫瘍がある
  - 表面は褐色調
  - 3x2cm大の潰瘍を有する
  - 肛門側断端からは11cm、口側断端からは2cm離れている
  - 各断端に及んでいない
  - 漿膜側は硬くなっているが、露出は見られない
  - 割面では白色充実性で、出血や壊死は僅か
- 下部大弯に3mm大の亜有茎性ポリープ
  - 肛門側断端から6cm
- 他の粘膜面に著変なし

検体に付いている

腫瘍の表面

腫瘍の漿膜側

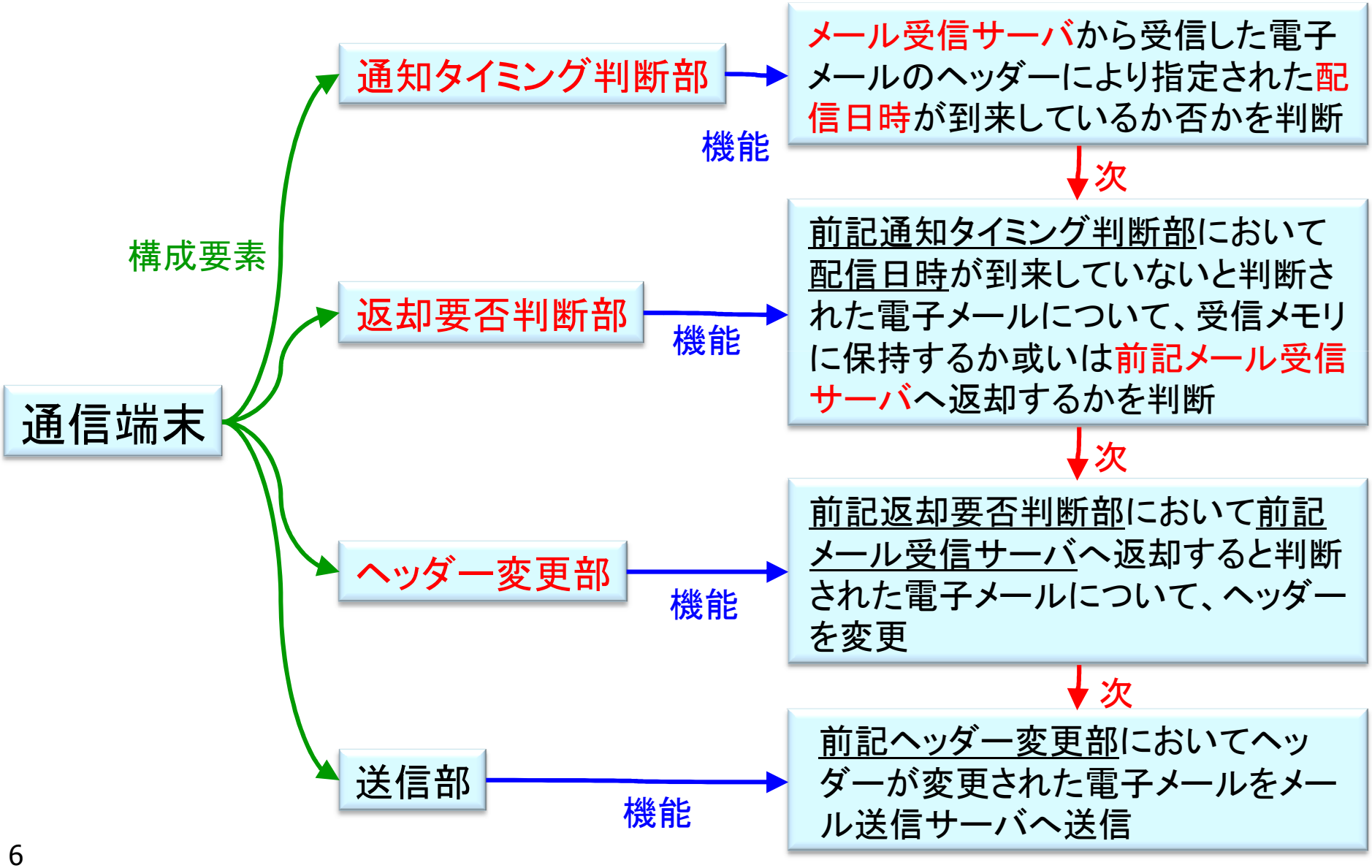
腫瘍が露出

検体の割面

上の腫瘍とここのポリープは別々のものであることが木構造からわかるので、「以上の主病変と別」である旨は記述不要

検体の肛門側断端

仮想特許の請求項 (グラフ形式)



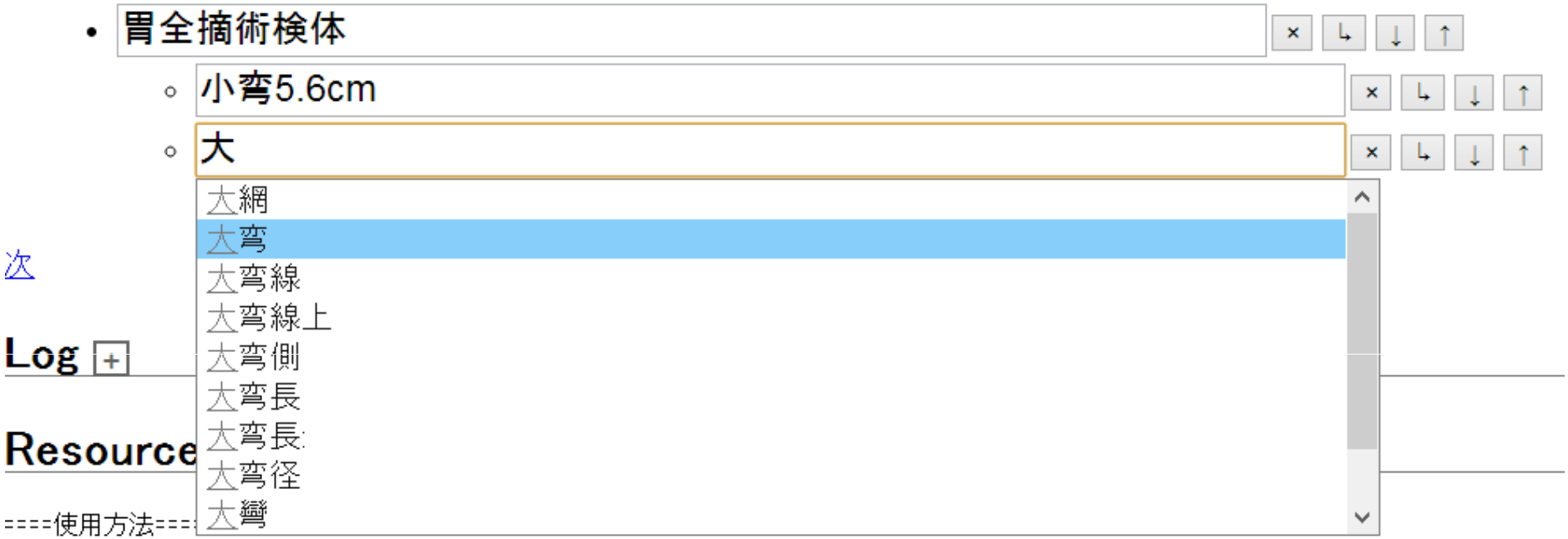
ノード同士の関係

- 空間的関係
  - 包含、付加、距離、他
- 時間的関係
  - 後続、因果、目的、機能、他
- 病理診断報告書等に含まれるほとんどの関係は空間的
- 特許の請求項等は時間的関係も含む

グラフ(木)形式の有効性の検証

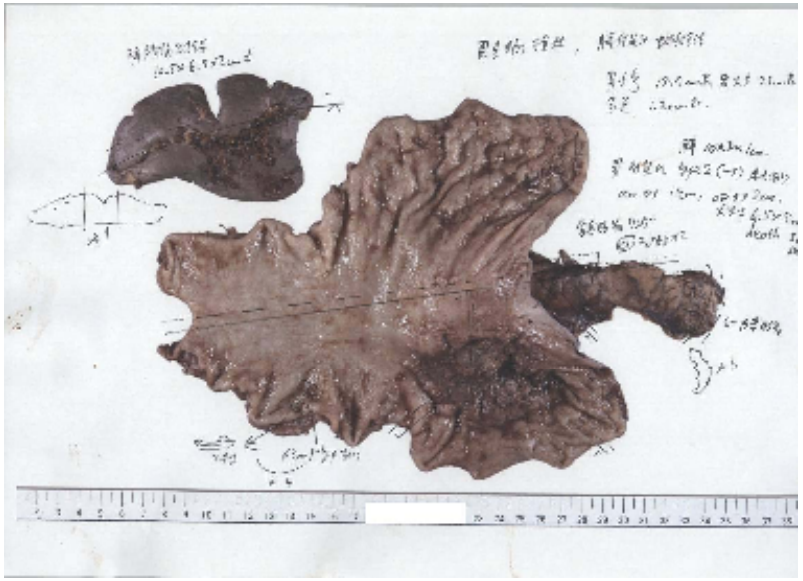
- 実験
  - 10人の病理医の各々が10個の症例(依頼書と写真)のうち5個を従来形式、残り5個をグラフ形式で記述。
  - わずか5分程度の教示により、グラフ形式を用いて、従来形式の場合と同等以上の品質と効率で病理診断報告を記述することができた。
- 考察
  - 習熟によってグラフ形式の方が従来形式よりも品質と効率が有意に高まる可能性がある。
  - グラフ形式に基づく精度の高い入力支援を提供することにより、従来よりはるかに高い品質と効率で文書を作成できると期待される。

木形式によるテキスト入力補完



- 木構造が規定する文脈に応じた処理
  - テキストの解析・予測・補完
  - 意味構造の生成

病理診断報告書(木形式)

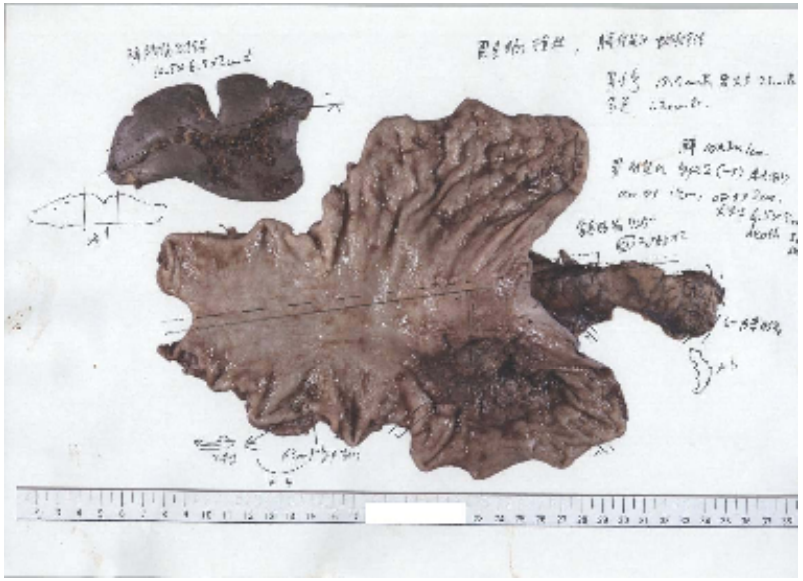


胃全摘検体

- 小弯長16cm、大弯長23cm
- 口側周径3cm、肛門側周径4.5cm
- 1.2cm長の食道が付いている
- 上部後壁に6.5x5cm大の2型腫瘍がある
  - 表面は褐色調
  - 3x2cm大の潰瘍を有する
  - 肛門側断端からは11cm、口側断端からは2cm離れている
  - 各断端に及んでいない
  - 漿膜側は硬くなっているが、露出は見られない
  - 割面では白色充実性で、出[illegible]や壊死は僅か
- 下部大弯に3m[illegible]プ
  - 肛門側断[illegible]
- 他の粘膜面に著変なし

直列的照応詞

病理診断報告書(木形式)



胃全摘検体

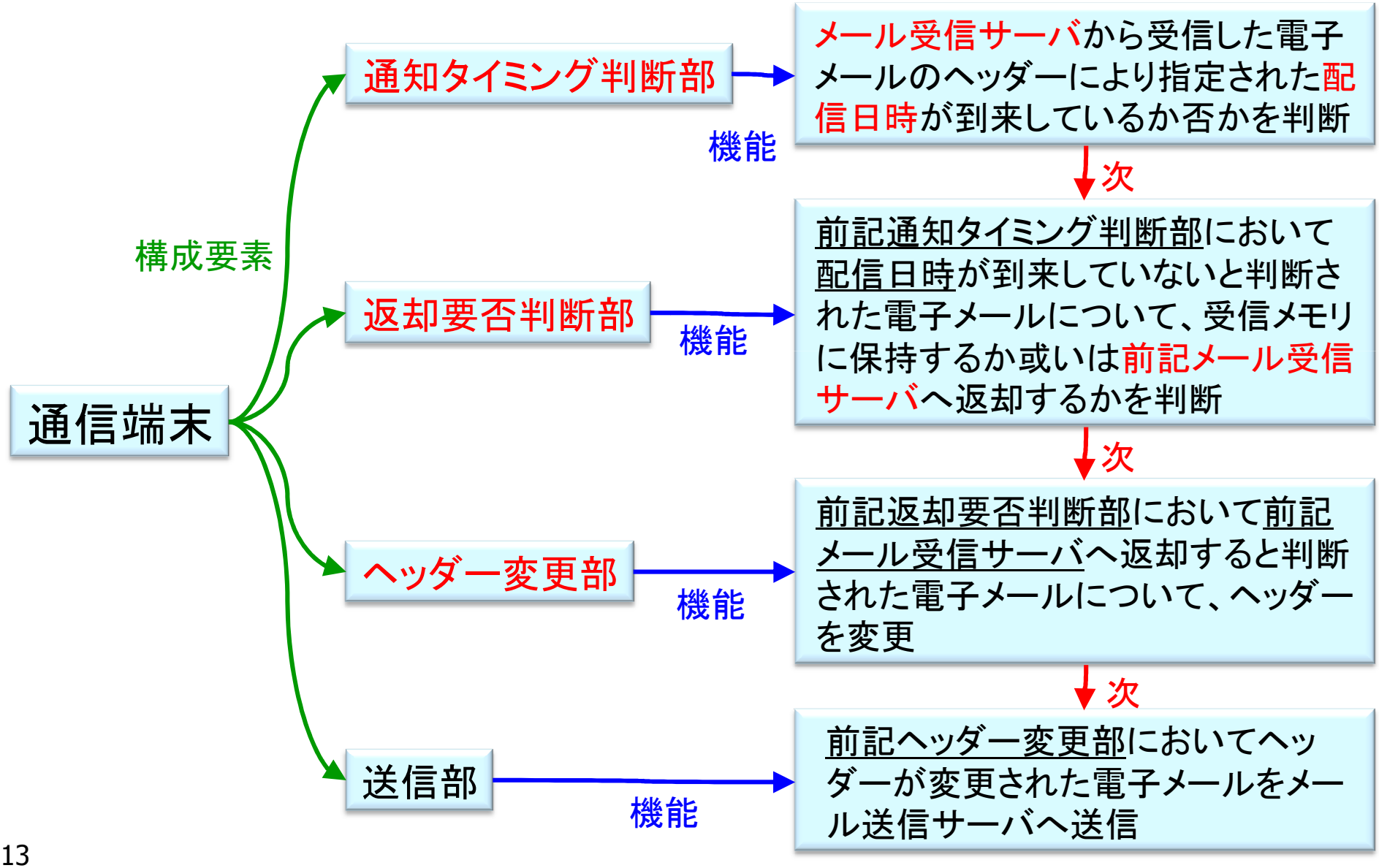
- 小弯長16cm、大弯長23cm
- 口側周径3cm、肛門側周径4.5cm
- 1.2cm長の食道が付いている
- 上部後壁に6.5x5cm大の2型腫瘍がある
  - 表面は褐色調
  - 3x2cm大の潰瘍を有する
  - 肛門側断端からは11cm、口側断端からは2cm離れている
  - 各断端に及んでいない
  - 漿膜側は硬くなっているが、露出は見られない
  - 割面では白色充実性で、出血や壊死は僅か
- 下部大弯に3mm大の亜有茎性ポリープ
  - 肛門側断端から6cm
- 他の粘膜面に著変なし

並列的照応詞

仮想特許の請求項 (従来形式)

- メール受信サーバから受信した電子メールのヘッダーにより指定された配信日時が到来しているか否かを判断する通知タイミング判断部と、
- 前記通知タイミング判断部において配信日時が到来していないと判断された電子メールについて、受信メモリに保持するか或いは前記メール受信サーバへ返却するかを判断する返却要否判断部と、
- 前記返却要否判断部において前記メール受信サーバへ返却すると判断された電子メールについて、ヘッダーを変更するヘッダー変更部と、
- 前記ヘッダー変更部においてヘッダーが変更された電子メールをメール送信サーバへ送信する送信部と、
- を備える通信端末。

仮想特許の請求項 (グラフ形式)



展望

- 照応詞と先行詞とのグラフ中の位置関係に関する制約の精密化
  - 人間が特に意識しなくても自然に満たされる制約
- 文法とオントロジーの漸進的拡張の手法の開発
  - 特許オントロジーの開発
- 利用者と目的の明確化
  - 意味構造の明示による一次・二次利用の価値向上
  - 専門家による文書作成を支援するのが目的ではない