Visual Tagging Tool (VTT)

Dr. Chris J. Lu
The Lexical Systems Group
NLM. LHNCBC. CGSB
http://SPECIALIST.nlm.nih.gov/vtt
12, 2010

<u>Table of Contents</u>

- Introduction

- <u>Manual tagging</u>

- <u>VTT file format (integration)</u>

- <u>Questions</u>

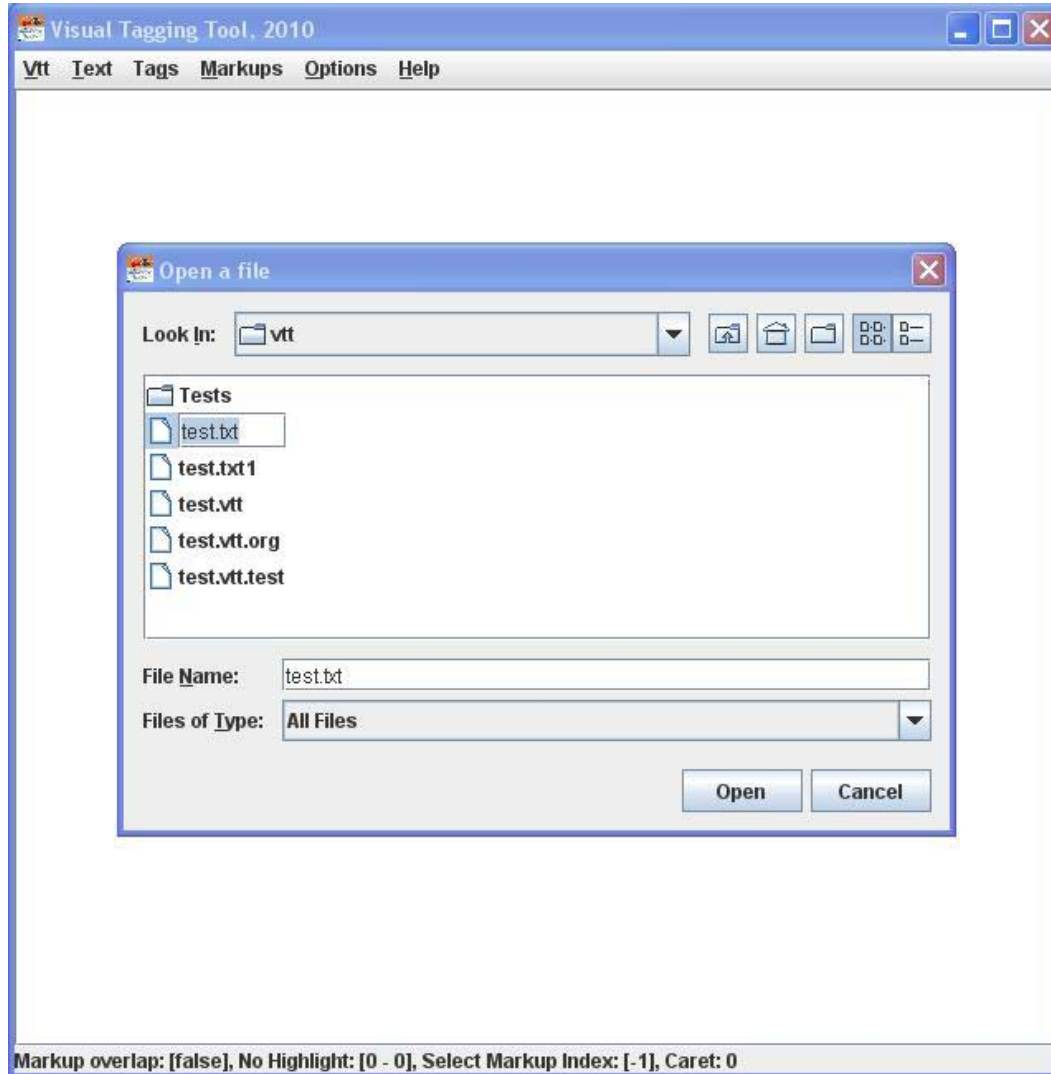# <u>Introduction</u>

- A simple, easy, lightweight, portable, Java Swing based annotation tool
- Shows tagged text in different visual effects: color, font, size, **bold**, *italic*, underline, etc.
- Developed to ease the human tagging process (markup text)
- Can be integrated with other NLP programs
- Full documents & supports
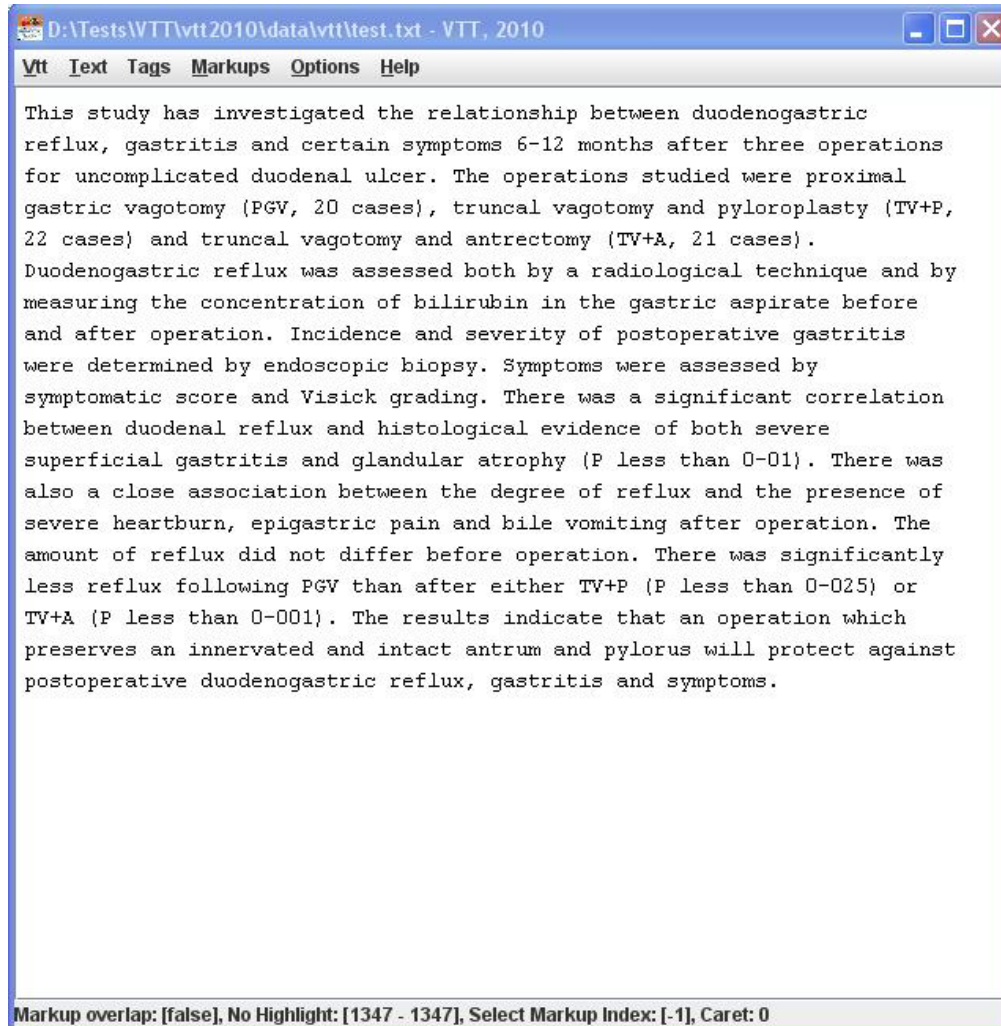- Free distributed with open source code (since 2009)

Basic Steps to Use VTT?

1) Open a plain unmarked text

- Vtt -> Open (O)

2) Define or import tags

- Tags -> Setup

- Tags -> Setup -> Import

3) Start to markup (assign tag to plain text)

- Select text: smear, double clicks, triple clicks, quick keys, etc.
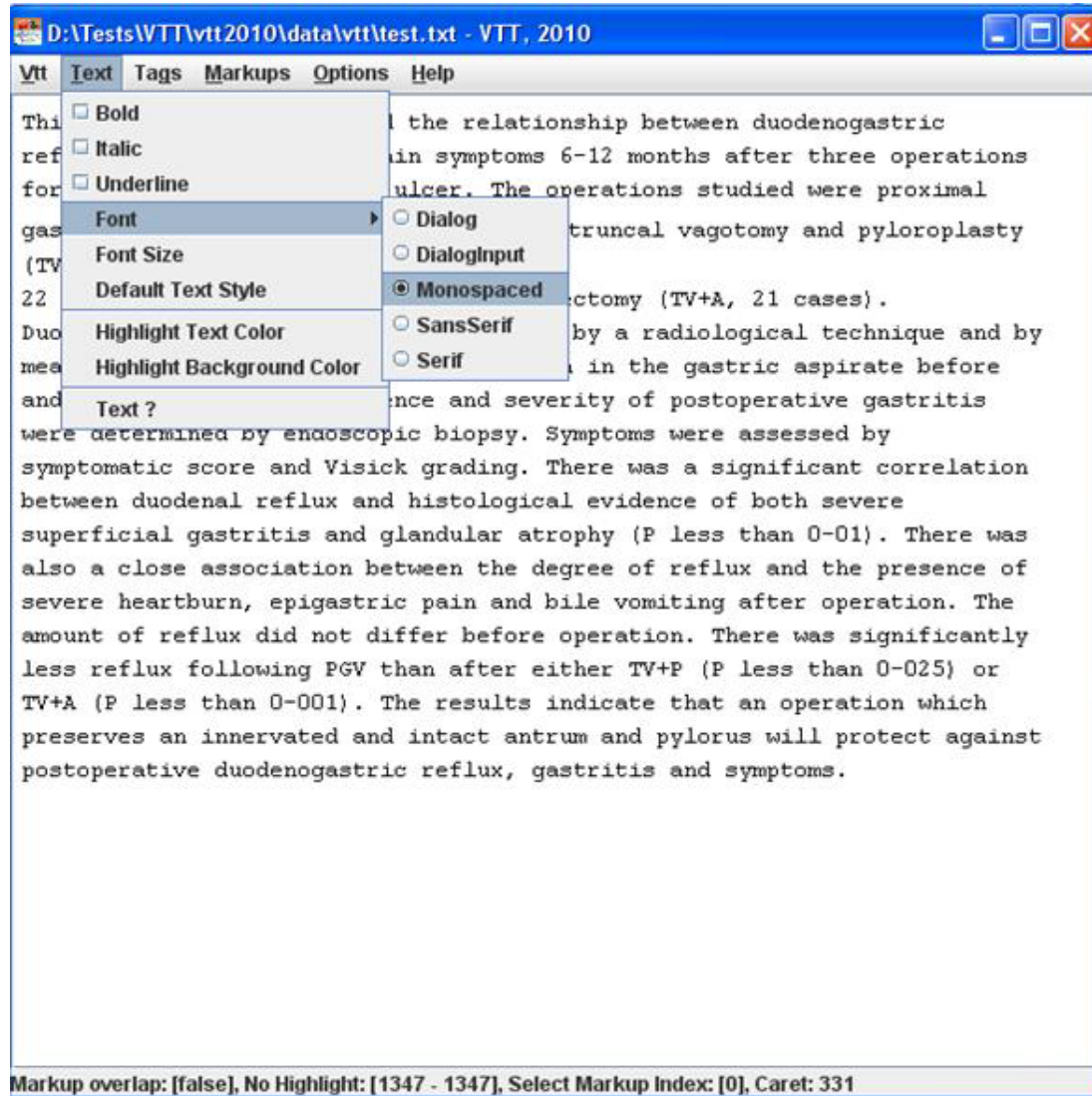
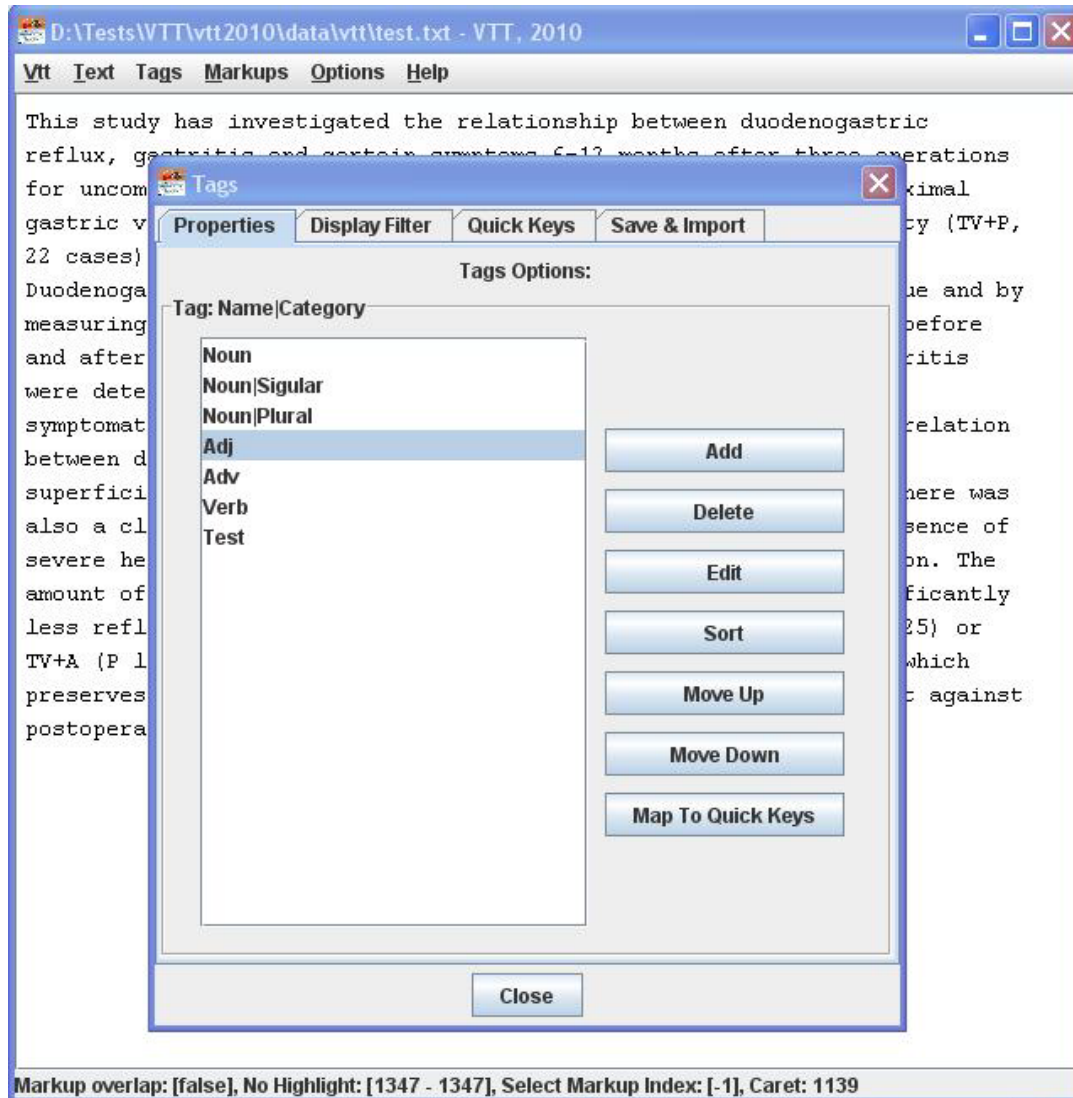- Assign tag: quick keys, pull-down menu
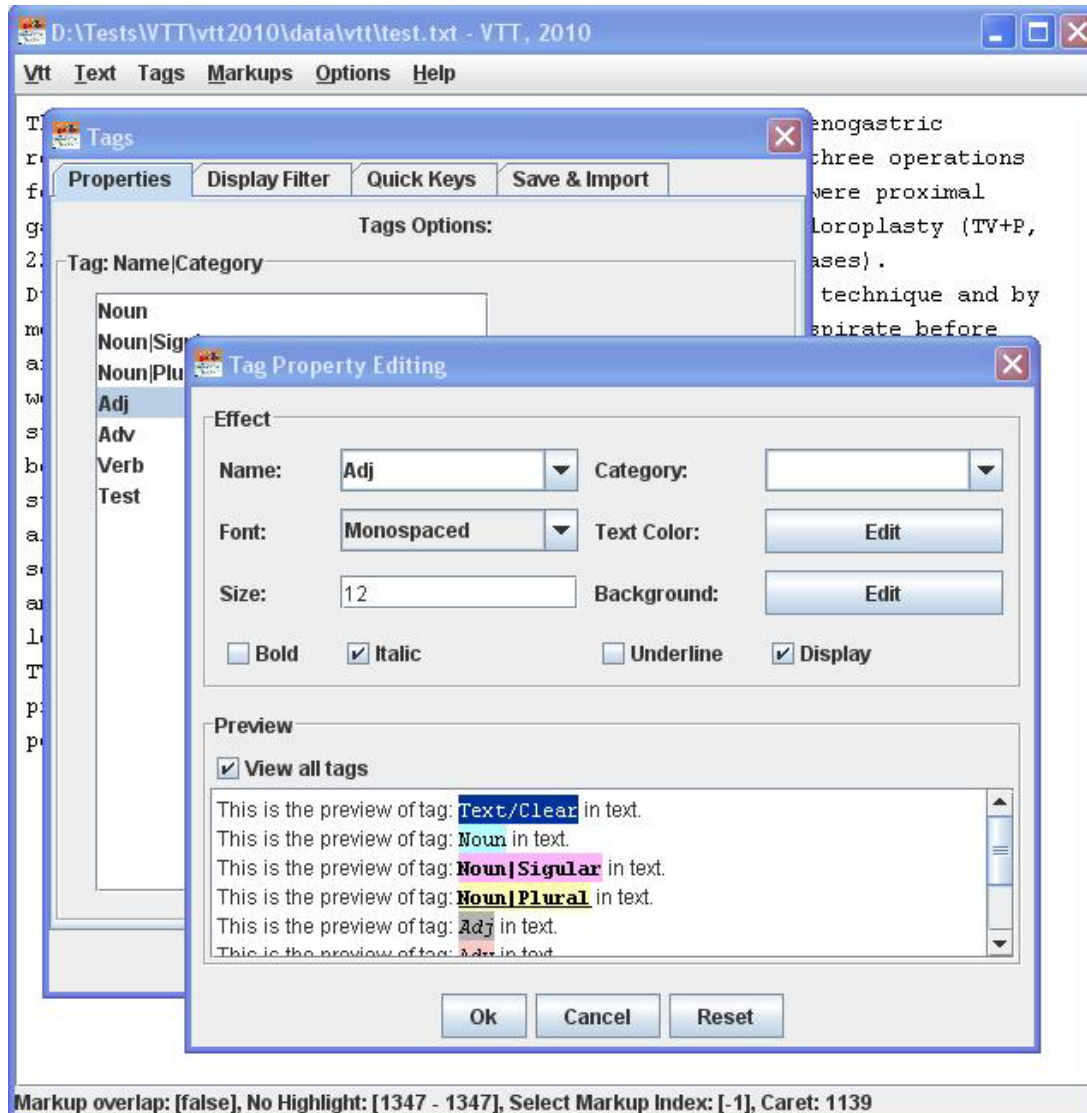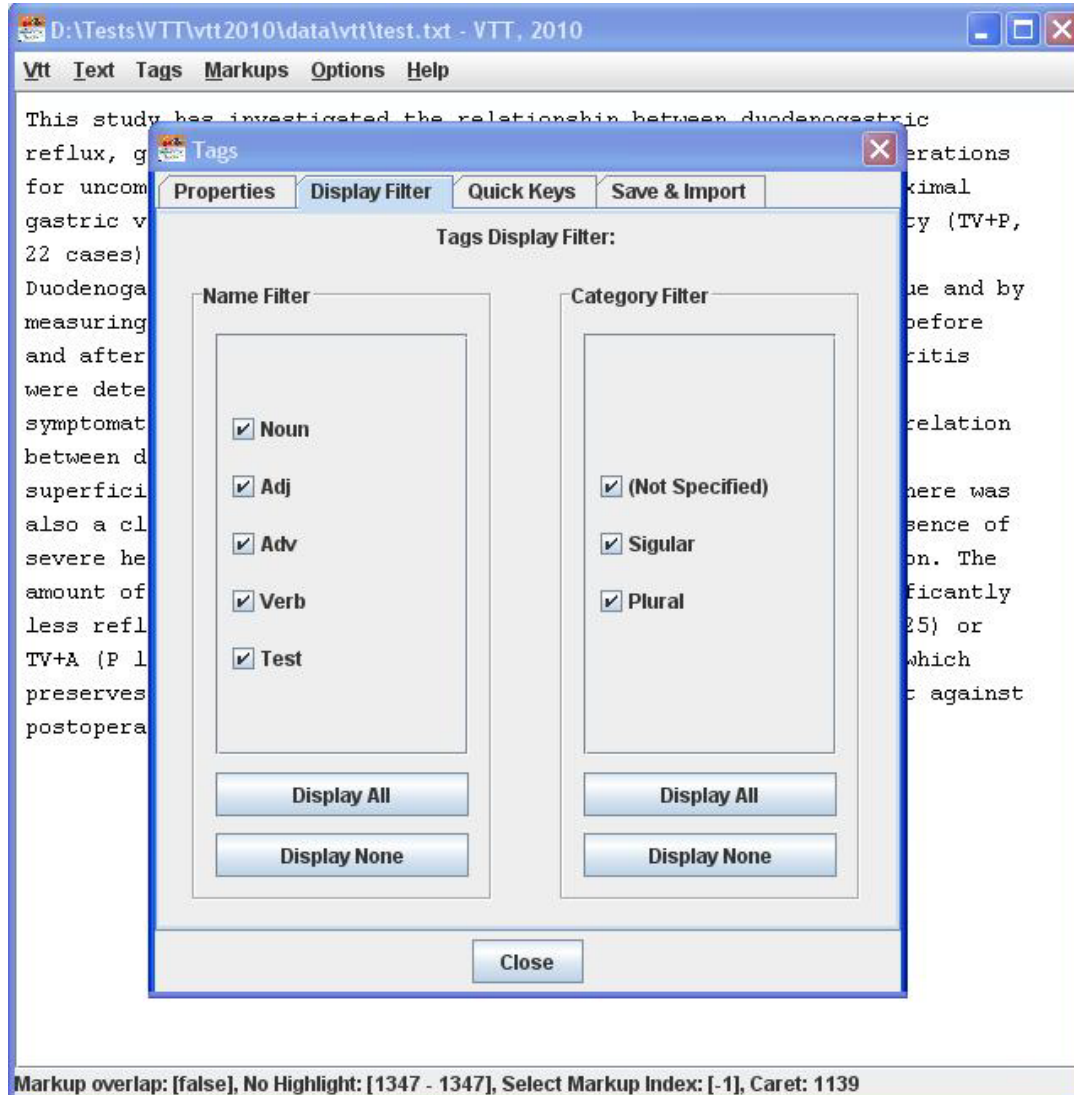
Text - Open

Text – Open (Cont.)

Text - Options

Tags - Setup

Tags - Edit

Tags – Display Filter

Tags - Quick Keys

Tags – Save & Import

Markups – Select Text
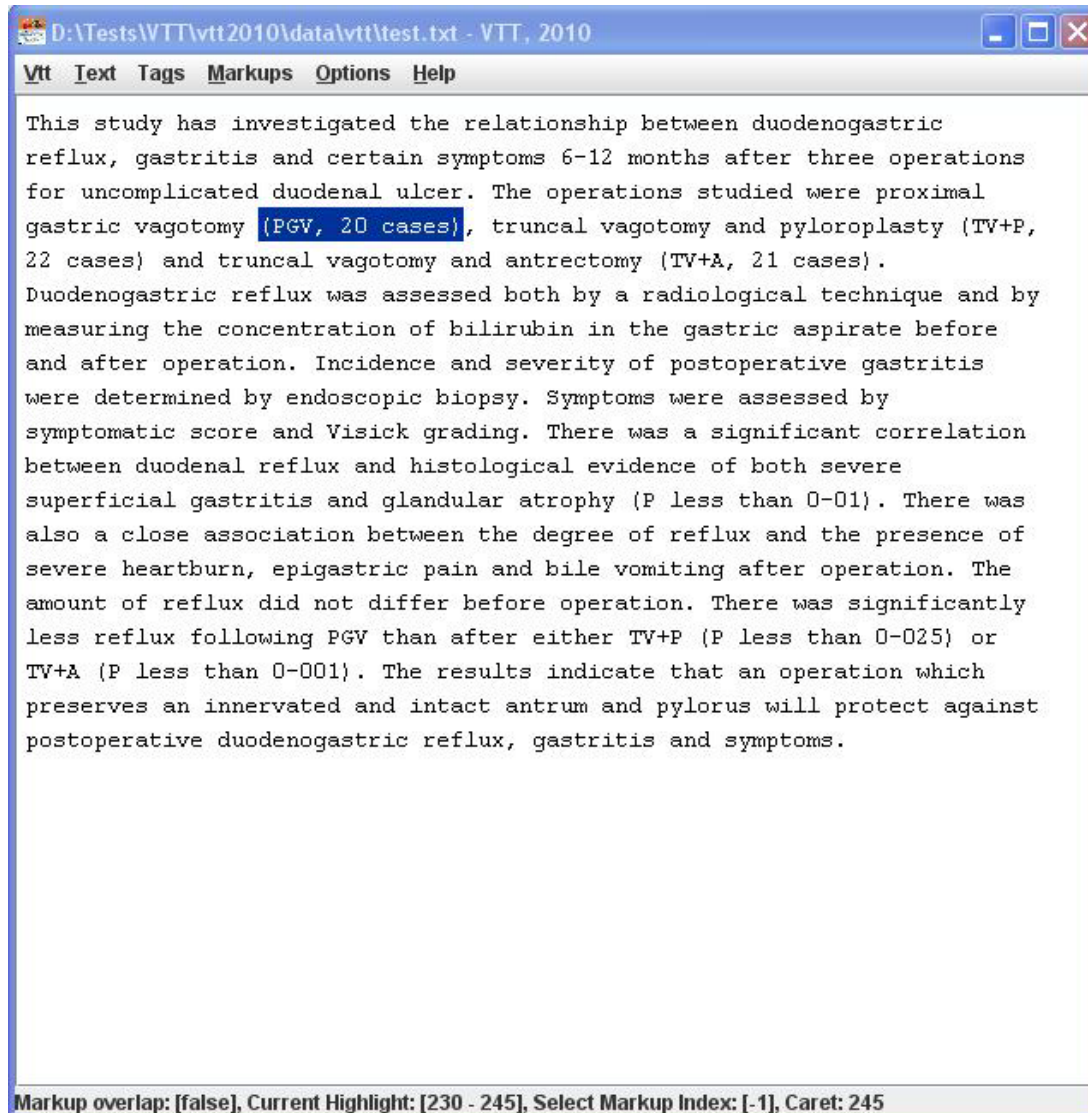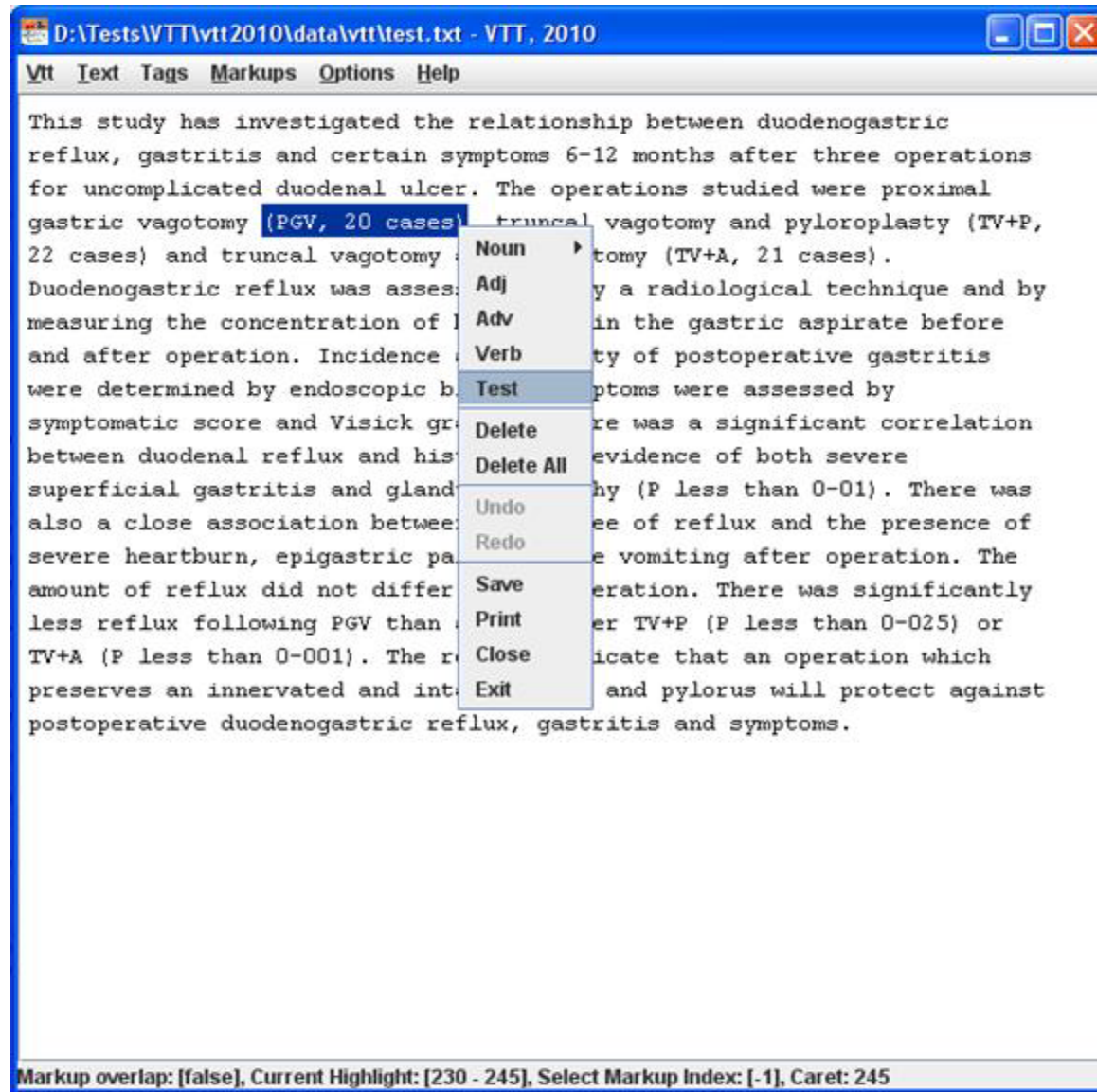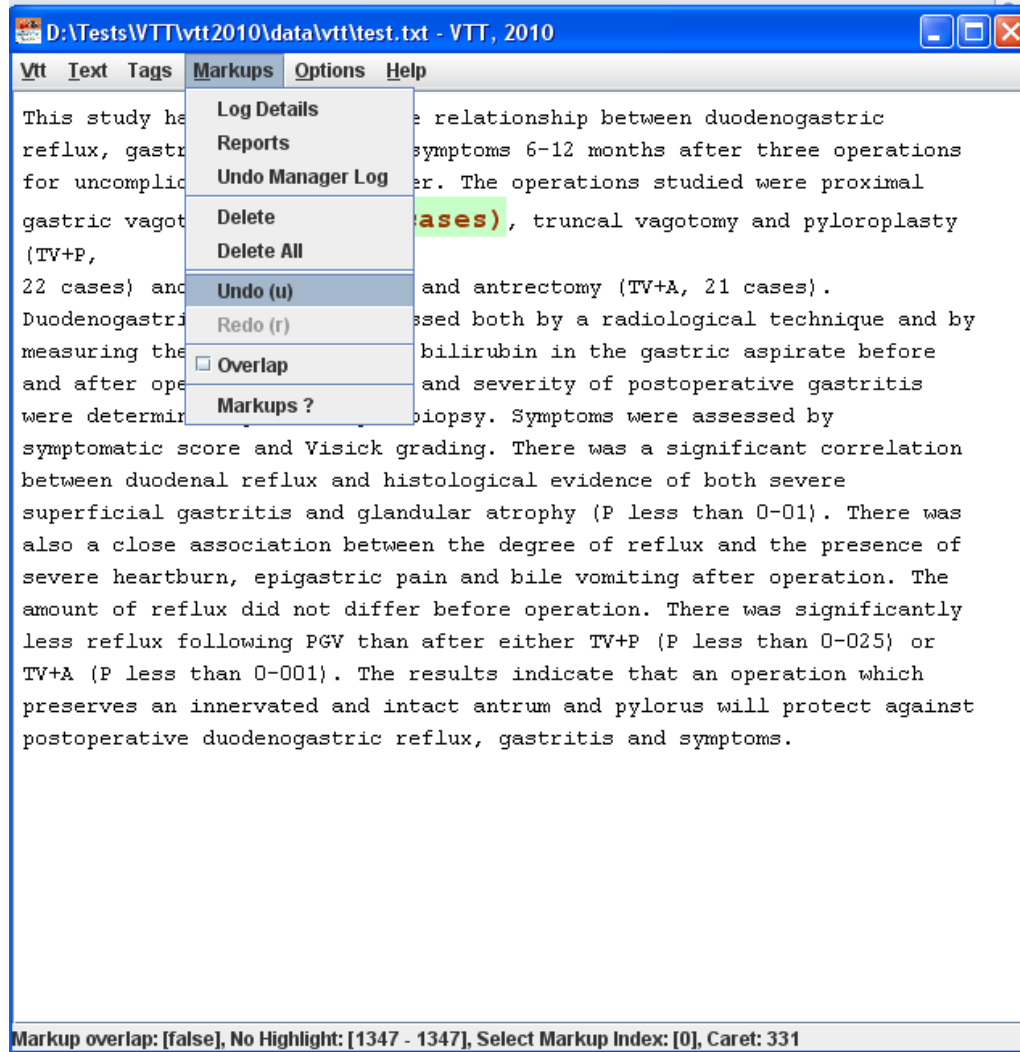
Markups - Assign a Tag

Markups - Assign a Tag
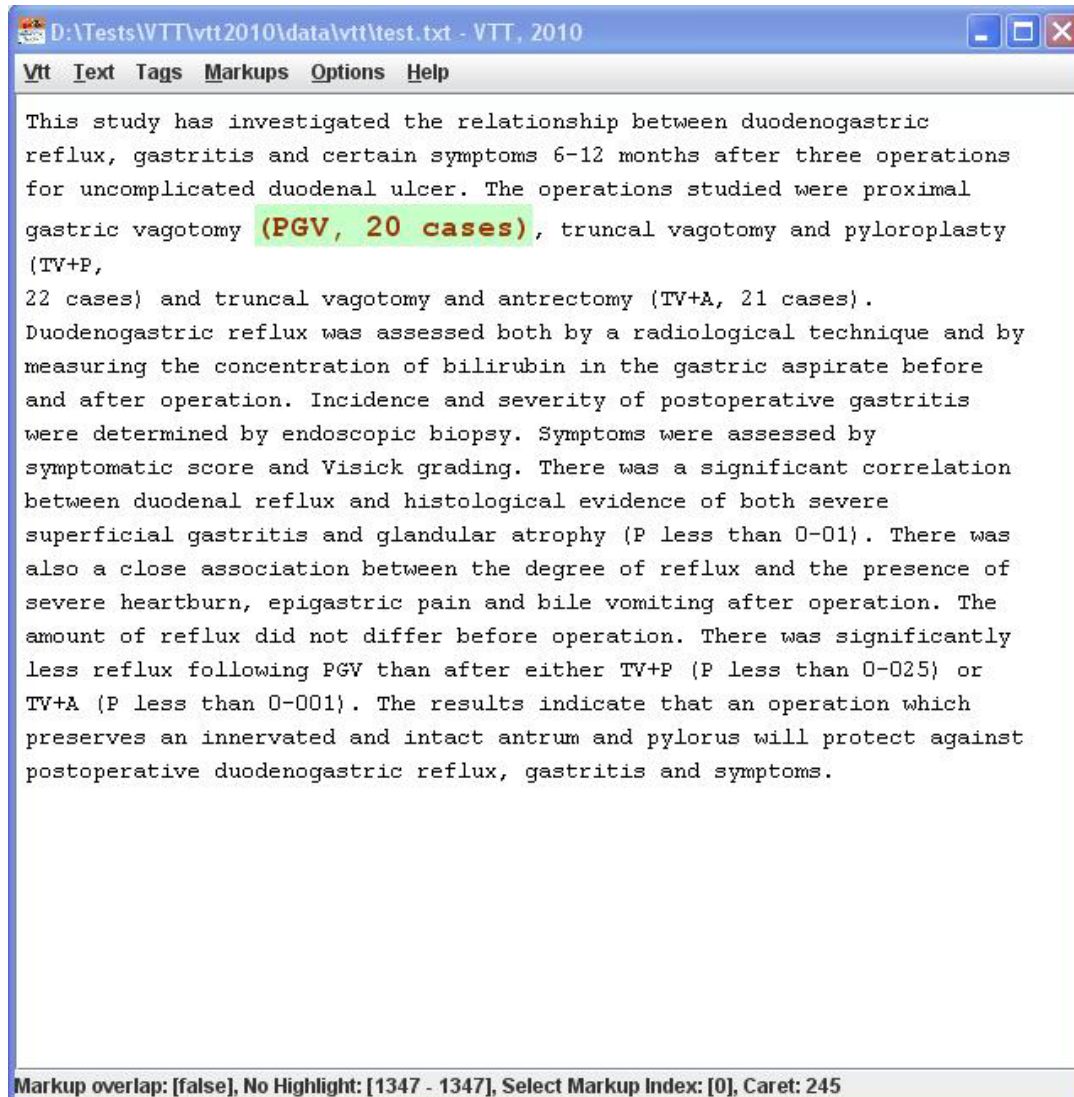
D:\Tests\VTT\vtt2010\data\vtt\test.txt - VTT, 2010

Vtt Text Tags Markups Options Help

This study has investigated the relationship between duodenogastric reflux, gastritis and certain symptoms 6-12 months after three operations for uncomplicated duodenal ulcer. The operations studied were proximal gastric vagotomy **(PGV, 20 cases)**, truncal vagotomy and pyloroplasty (TV+P,
22 cases) and truncal vagotomy and antrectomy (TV+A, 21 cases). Duodenogastric reflux was assessed both by a radiological technique and by measuring the concentration of bilirubin in the gastric aspirate before and after operation. Incidence and severity of postoperative gastritis were determined by endoscopic biopsy. Symptoms were assessed by symptomatic score and Visick grading. There was a significant correlation between duodenal reflux and histological evidence of both severe superficial gastritis and glandular atrophy (P less than O-O1). There was also a close association between the degree of reflux and the presence of severe heartburn, epigastric pain and bile vomiting after operation. The amount of reflux did not differ before operation. There was significantly less reflux following PGV than after either TV+P (P less than O-O25) or TV+A (P less than O-OO1). The results indicate that an operation which preserves an innervated and intact antrum and pylorus will protect against postoperative duodenogastric reflux, gastritis and symptoms.

Markup overlap: [false], No Highlight: [1347 - 1347], Select Markup Index: [0], Caret: 245

Markups - Options

Markups – Logs

Markups – Reports

Markups – More Features

- Delete selected markup (0)
- Assign a tag by quick keys (1~9)
- Markup redo (r) /undo (u)
- Markups join (j)
- Markups movement control by keys
- Selected Markup Information (m)

Markups – Informations

Markups – Movement

- **Markups related**

Key	Action
[	Set selected markup to the first markup
]	Set selected markup to the last markup
,	Backward selected markup
.	Forward selected markup
{	Set selected markup to the first displayed markup
}	Set selected markup to the last displayed markup
<	Backward selected displayed markup
>	Forward selected displayed markup
j	Join the selected markup and the next markup
m	Shows markup detail dialog on the selected markup
r	Redo markup operation
u	Undo markup operation
0	Delete selected markup
1 ~ 9	Set selected markup/highlight text to tag (1 ~ 9)

Other Options
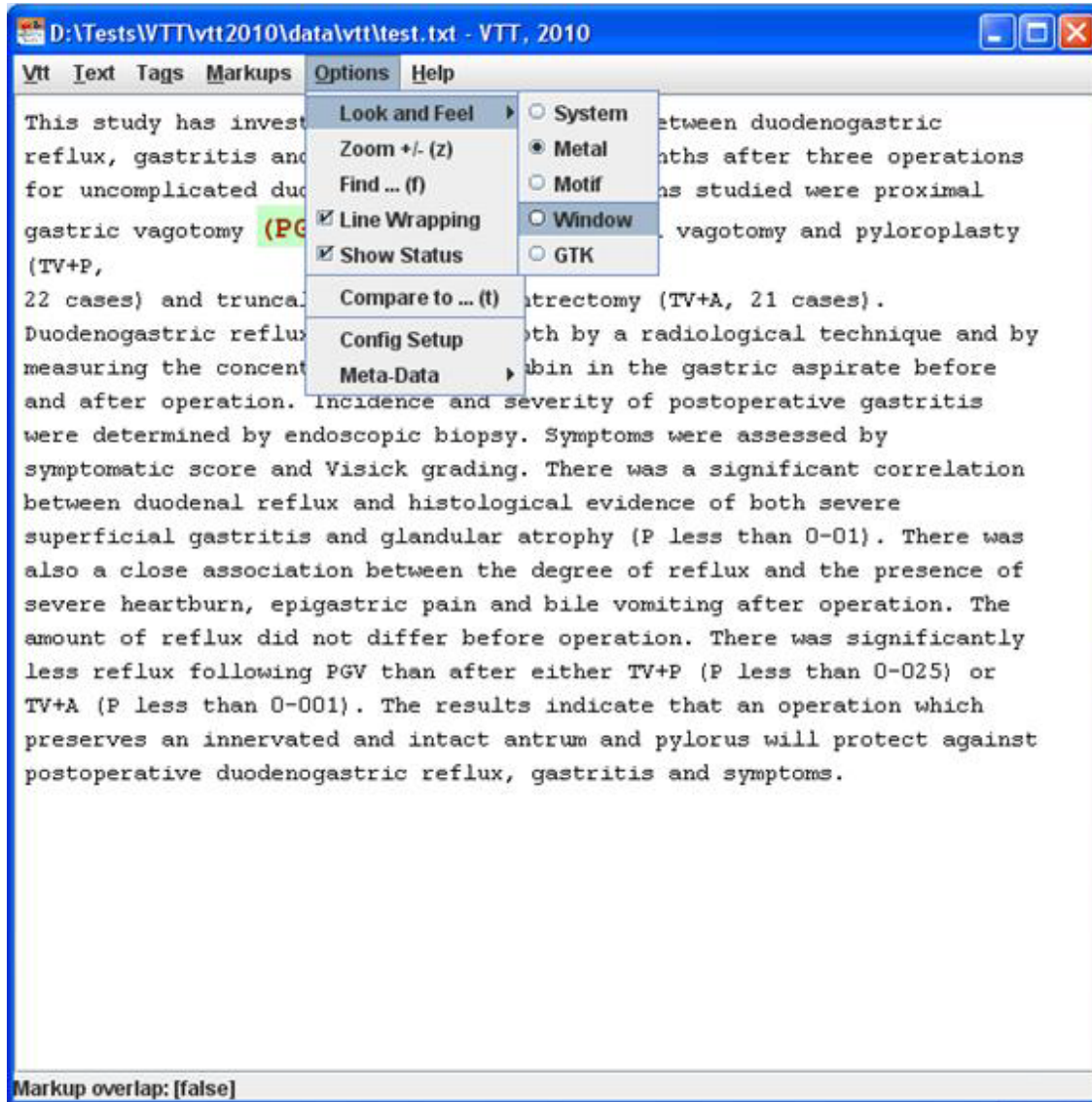
Compare Option

VTT Compare: Different Markups from Source and Target Files

View Markups Diff | View Data Diff | Stats Reports | Difference | Raw Data
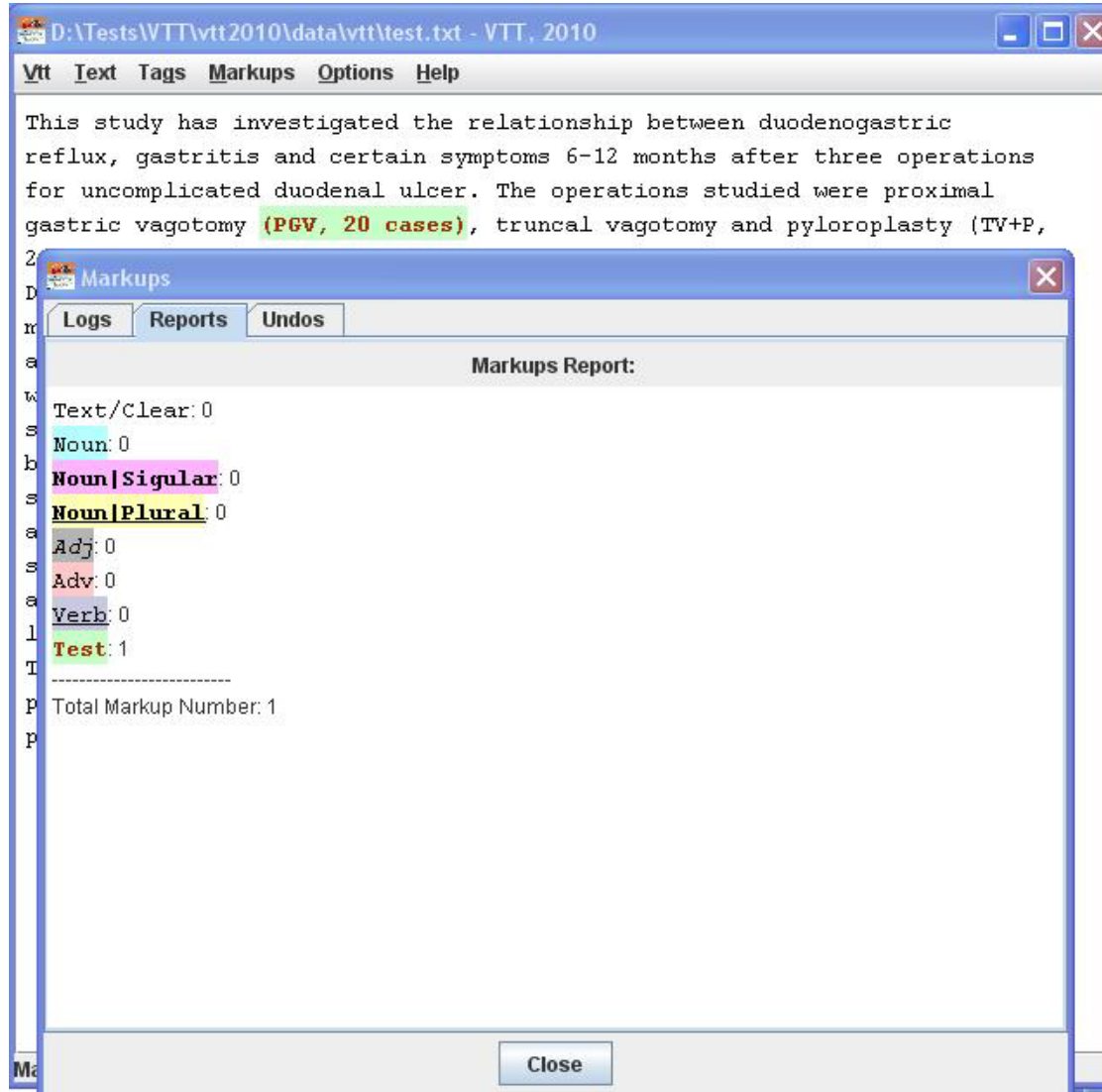
Markups in Source File

This study has investigated the relationship between duodenogastric reflux, gastritis and certain symptoms 6-12 months after three operations for uncomplicated duodenal ulcer. The operations studied were proximal gastric vagotomy (PGV, 20 cases), truncal vagotomy and pyloroplasty (TV+P, 22 cases) and truncal vagotomy and antrectomy (TV+A, 21 cases). Duodenogastric reflux was assessed both by a radiological technique and by measuring the concentration of bilirubin in the gastric aspirate before and after operation. Incidence and severity of postoperative gastritis were determined by endoscopic biopsy. Symptoms were assessed by symptomatic score and Visick grading. There was a significant correlation between duodenal reflux and histological evidence of both severe superficial gastritis and glandular atrophy (P less than 0-01). There was also a close association between the degree of reflux and the presence of severe heartburn, epigastric pain and bile vomiting after operation. The amount of reflux did not differ before operation. There was significantly less reflux following PGV than after either TV+P (P less than 0-025) or TV+A (P less than 0-001). The results indicate that an operation which preserves an innervated and intact antrum and pylorus will protect against postoperative duodenogastric reflux, gastritis and symptoms.

Markups in Target File

This study has investigated the relationship between duodenogastric reflux, gastritis and certain symptoms 6-12 months after three operations for uncomplicated duodenal ulcer. The operations studied were proximal gastric vagotomy (PGV, 20 cases), truncal vagotomy and pyloroplasty (TV+P, 22 cases) and truncal vagotomy and antrectomy (TV+A, 21 cases). Duodenogastric reflux was assessed both by a radiological technique and by measuring the concentration of bilirubin in the gastric aspirate before and after operation. Incidence and severity of postoperative gastritis were determined by endoscopic biopsy. Symptoms were assessed by symptomatic score and Visick grading. There was a significant correlation between duodenal reflux and histological evidence of both severe superficial gastritis and glandular atrophy (P less than 0-01). There was also a close association between the degree of reflux and the presence of severe heartburn, epigastric pain and bile vomiting after operation. The amount of reflux did not differ before operation. There was significantly less reflux following PGV than after either TV+P (P less than 0-025) or TV+A (P less than 0-001). The results indicate that an operation which preserves an innervated and intact antrum and pylorus will protect against postoperative duodenogastric reflux, gastritis and symptoms.

Ok | Open Target File

VTT – File Format

- **Meta Data**
  - Default tag file (3 fields)
  - History (4 fields)
- **Text Content**
  - Original plain text
- **Tags Configuration (14 fields)**
  - Name and category
  - Bold, italic, underline, display, font family, font size
  - Foreground and background colors (RGB)
- **Markups Information (6 fields)**
  - Offset and length
  - Assign tag (name and category)
  - Annotation and tagged text

File Format – Meta Data


```
test.vtt - Notepad
File  Edit  Format  View  Help
#<------------------------------------------------------------------>
#<Meta Data>
#<Tags file|confirmation|path of default tags file>
#<File History|VTT file format version|User name|Time stamp>
#<------------------------------------------------------------------>
TAGS_FILE|true|/export/home/lu/Development/VTT/vtt2010/data/config/tags.data
FILE_SAVE|VTT.2010.0|VTT_Guest|2/2/10 4:44:39 PM
FILE_SAVE|VTT.2010.0|chlu|2/2/10 4:45:44 PM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/2/10 4:46:14 PM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/3/10 4:36:29 PM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/3/10 4:43:39 PM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/4/10 11:57:26 AM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/4/10 11:57:27 AM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/4/10 11:57:28 AM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/4/10 11:57:29 AM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/4/10 11:57:30 AM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/4/10 11:57:42 AM
FILE_SAVE|VTT.2010.0|VTT_Guest|2/4/10 11:57:43 AM
```

File Format – Text


```
#<-------------------------------------------------------------------->
#<Text Content>
#<-------------------------------------------------------------------->
This study has investigated the relationship between duodenogastric
reflux, gastritis and certain symptoms 6-12 months after three operations
for uncomplicated duodenal ulcer. The operations studied were proximal
gastric vagotomy (PGV, 20 cases), truncal vagotomy and pyloroplasty (TV+P,
22 cases) and truncal vagotomy and antrectomy (TV+A, 21 cases).
Duodenogastric reflux was assessed both by a radiological technique and by
measuring the concentration of bilirubin in the gastric aspirate before
and after operation. Incidence and severity of postoperative gastritis
were determined by endoscopic biopsy. Symptoms were assessed by
symptomatic score and Visick grading. There was a significant correlation
between duodenal reflux and histological evidence of both severe
superficial gastritis and glandular atrophy (P less than 0-01). There was
also a close association between the degree of reflux and the presence of
severe heartburn, epigastric pain and bile vomiting after operation. The
amount of reflux did not differ before operation. There was significantly
less reflux following PGV than after either TV+P (P less than 0-025) or
TV+A (P less than 0-001). The results indicate that an operation which
preserves an innervated and intact antrum and pylorus will protect against
postoperative duodenogastric reflux, gastritis and symptoms.
#<-------------------------------------------------------------------->
```
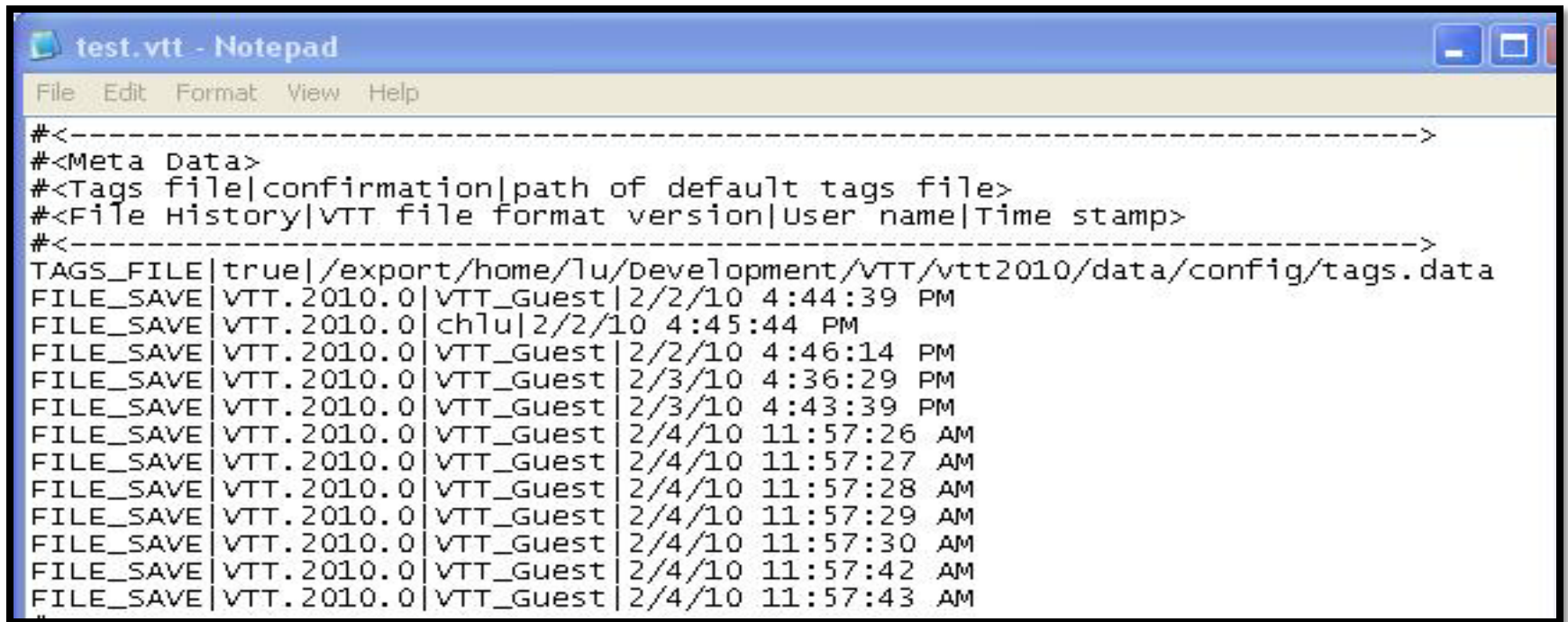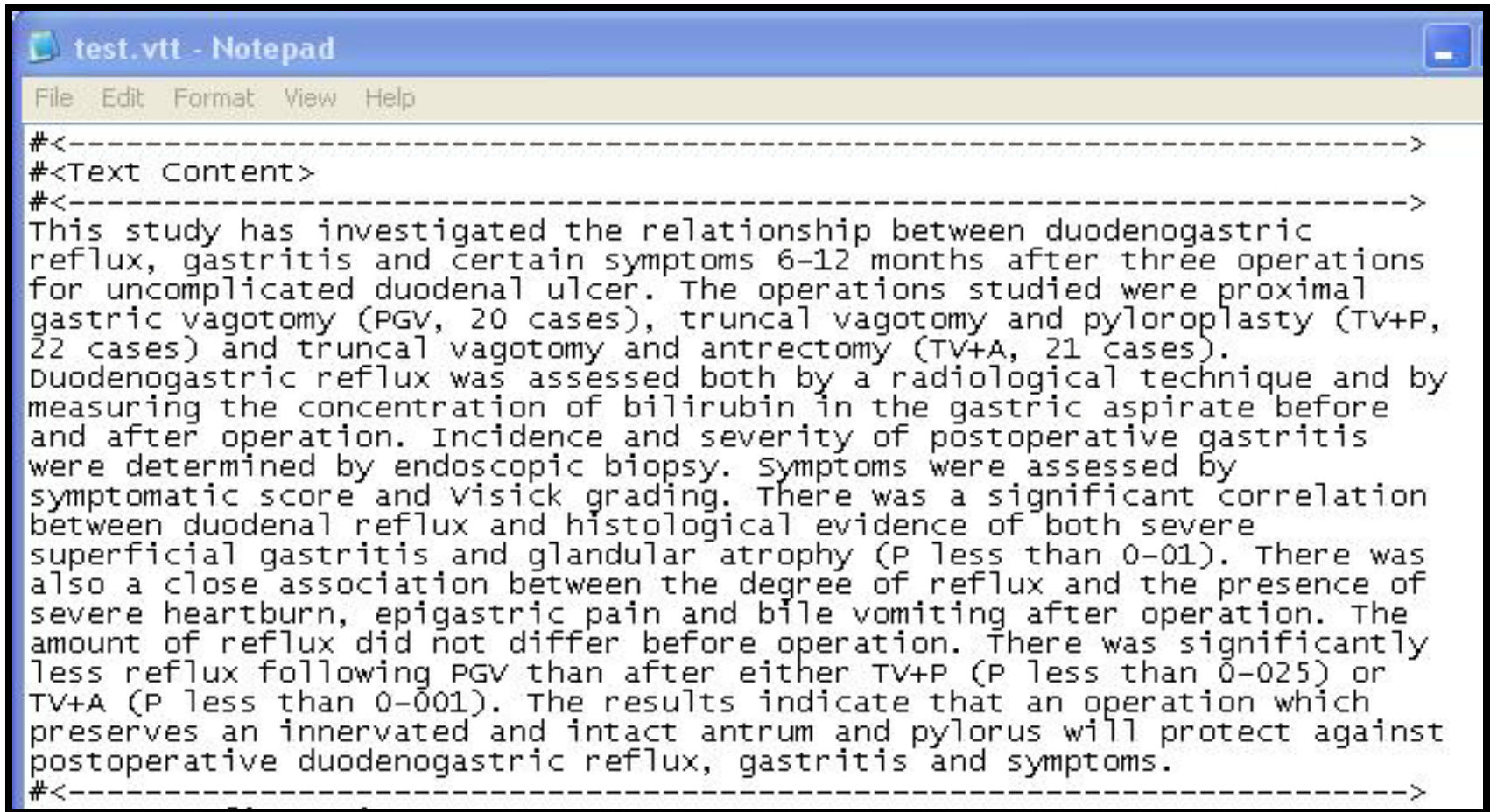
File Format – Tags


```
#<----------------------------------------------------->
#<Tags Configuration>
#<Name|Category|Bold|Italic|Underline|Display|FR|FG|FB|BR|BG|BB|FontFamily|Fon
tSize>
#<----------------------------------------------------->
Text/Clear||false|false|false|true|255|255|255|0|51|153|Monospaced|12
#<----------------------------------------------------->
Noun||false|false|false|true|0|0|0|255|180|180|Monospaced|12
Noun|Sigular|true|false|false|true|0|0|0|255|180|180|Monospaced|12
Noun|Plural|true|false|true|true|0|0|0|255|180|180|Monospaced|12
Adj||false|true|false|true|0|0|0|180|225|180|Monospaced|12
Adv||false|false|false|true|0|0|0|204|204|255|Monospaced|12
Verb||false|false|true|true|0|0|0|150|196|225|Monospaced|12
Test||true|false|false|true|153|51|0|255|255|204|Monospaced|12
#<----------------------------------------------------->
```

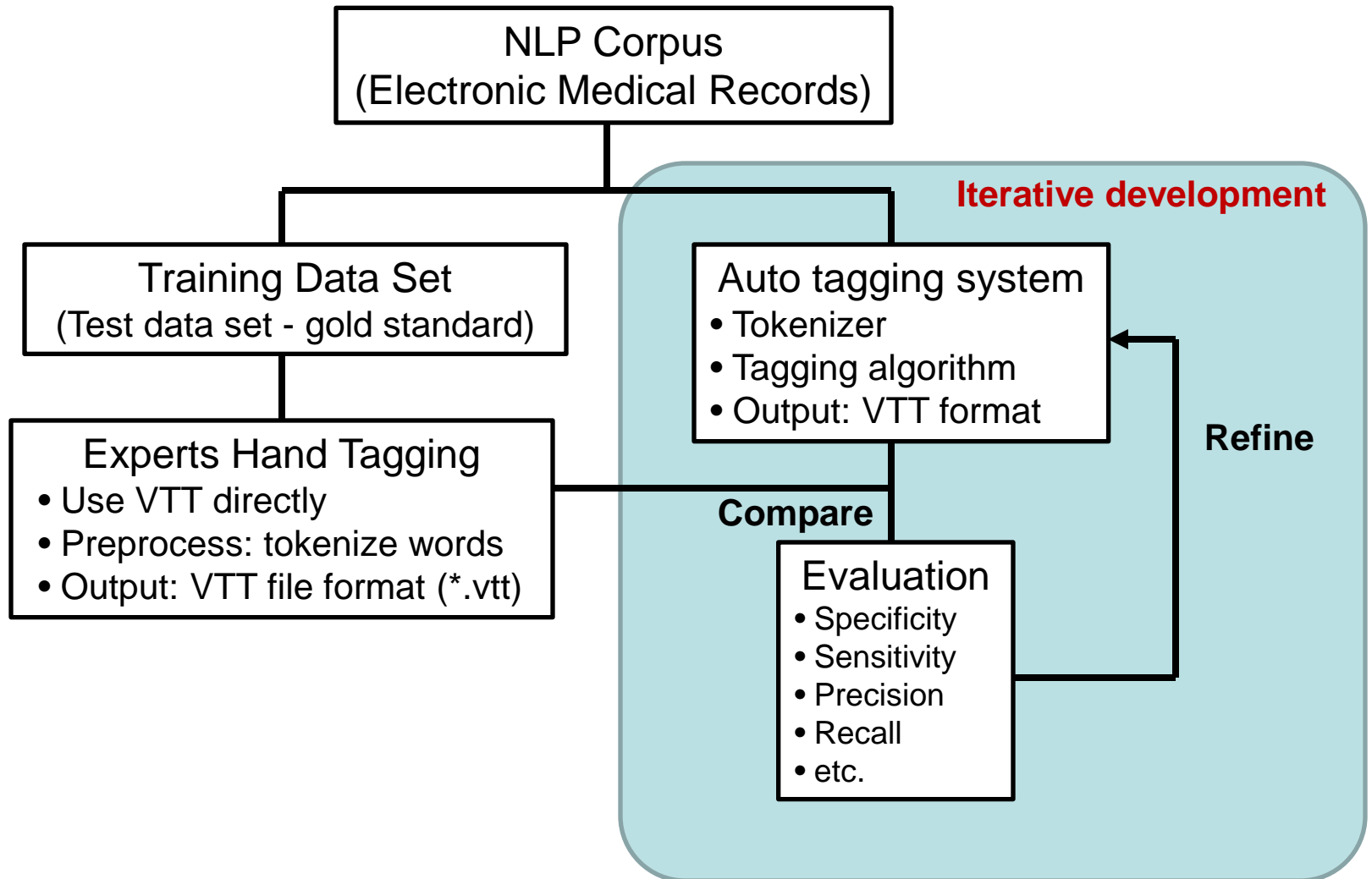# File Format - Markups


```
test.vtt - Notepad

File  Edit  Format  View  Help

#<------------------------------------------------------------------>
#<MarkUps Information>
#<Offset|Length|TagName|TagCategory|Annotation|TagText>
#<------------------------------------------------------------------>
       0|  4|Test         |            |First Test|This
       5|  5|Noun         |Sigular     ||study
      90|  7|Adj          |            ||certain
     112|  6|Adv          |            ||months
     247|  6|Verb         |            ||trunca
     367|  6|Adv          |            ||reflux
     373|  1|Noun         |Sigular     ||
     374|  3|Adv          |            ||was
     377|  1|Noun         |            ||
     378|  8|Test         |            ||assessed
     387|  4|Adv          |            ||both
     392|  2|Noun         |Plural      ||by
     397|12|Adj           |            ||radiological
     509|12|Adj           |            ||operation. I
     519|  2|Text/Clear   |            || I
     600|  7|Adv          |            |Extra space and S, need to
delete|biopsy.
     617|  4|Noun         |Sigular     |Were is a be verb|were
     736|12|Noun          |Sigular     ||histological
     894|  6|Noun          |Plural      ||reflux
     939|10|Noun          |            ||epigastric
     955|  3|Verb          |            ||and
    1337|  9|Adj           |            |Last test|symptoms.
```

VTT – Example

NLP Corpus
(Electronic Medical Records)

Iterative development

Training Data Set
(Test data set - gold standard)

Auto tagging system
- Tokenizer
- Tagging algorithm
- Output: VTT format

Refine

Experts Hand Tagging
- Use VTT directly
- Preprocess: tokenize words
- Output: VTT file format (*.vtt)

Compare
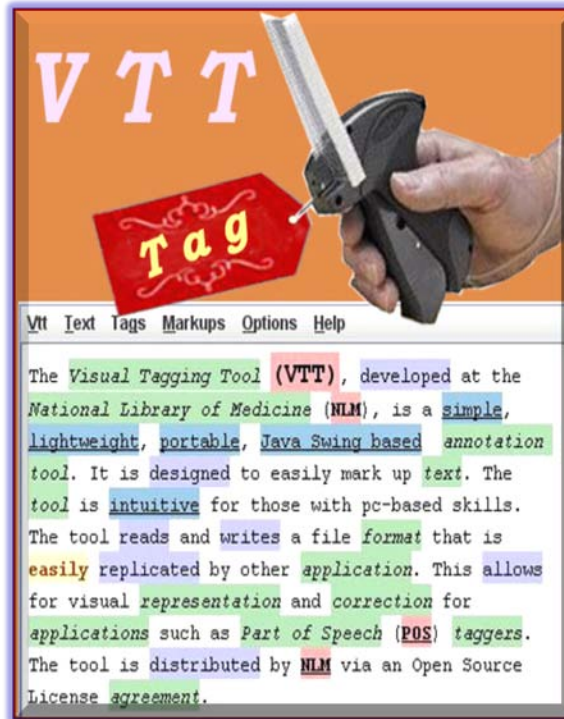
Evaluation
- Specificity
- Sensitivity
- Precision
- Recall
- etc.

VTT – Integration

- Use VTT file format for the NLP applications
- Integrate with words/terms tokenizer (text tools) to generate pre-tokenized VTT files for less key strokes
- Read in VTT files and apply evaluation software:
    - File level
    - Corpus level

Visual Tagging Tool (VTT)

http://SPECIALIST.nlm.nih.gov/vtt

Questions

- Lexical Systems Group: http://umlslex.nlm.nih.gov
- The SPECIALIST NLP Tools: http://specialist.nlm.nih.gov