

# **UW Internal Annotation Guidelines for the 2009 i2b2 Contest and UW Medication IE System**

Authors: Fei Xia, Imre Solti, and Ozlem Uzuner

Contributors: All other contributors on the Wiki

Date: 5/03/2009

**Purpose:** This document specifies the annotation guidelines for marking medications and related information in medical discharge summary documents. The annotated files will later be used to create the gold standard for the 2009 i2b2 contest and to train/test our IE systems.

## **Section 1: Overview**

### **1.1 Files**

We are dealing with four kinds of files:

- Discharge summary file: This is the input file for our annotation task. The file consists of free text.
- Annotated file: This is the output file for our annotation task. It is the same as discharged summary file except that sgml markups are added to indicate the location of various fields.
- Our internal entry file: The format is one entry per line. The fields in the same entry have the same index.
- The final entry file: This file has the same format as “our internal list file”, but its content can differ slightly to be consistent with i2b2 requirements.

The goal of this document is to answer questions for creating annotated files from the discharge summary files. We will say little about the internal and final entry files, other than they are generated automatically from the annotated files. For more information on the final entry files, please refer to the other guidelines written by the i2b2 challenge organizers.

### **1.2 Annotation tools**

Two annotation tools are provided for annotation:

- Callisto: Adapted by Andy Freeman
- GATE: Adapted by Joshua Johnson

You can use either tool for annotation. If you have any questions about the tools, please contact the authors of the plugins.

### **1.3 Fields**

We are marking the following seven **fields** in the *annotated* file:

- Medication name
- Dose
- Mode
- Frequency
- Reason
- Modifier
- Duration

## 1.4 Attributes

Attributes are name-value pairs. A field has multiple attributes.

All the fields have the following four **attributes**. The values for the first two must be set manually. The values for the last two are the same as the default values most of the time.

- tag: It indicates the type of fields. There are seven possible values, one for each field.
- index: Fields with the same index form an entry in the entry file. The value is an integer or a list of integers separated by commas.
- discontinuous: If the text for a field is separate by other words (e.g., “eye drop” in “eye and ear drop”), indicate this by setting the flag to 1. In all other cases, leave the flag to be 0 (the default value).
- antecedent: if the current field is a pronoun or a word such as “medication” that refers to medication names etc., the value of the attribute will be the index of its antecedent. In all other cases, the value for the attribute will be -1 (the default value).

The following attributes are available only for some fields:

- event, tense, certainty: it indicates the three dimensions of a “modifier”. They are used only for the modifier field. More detail in Section 2.6.
- section: It indicates whether the field is inside a list of medications or not. The value is “list” or “narrative”. The attribute is for the medication name only.
- output: it indicates whether a medication should be included in the final (i2b2) entry file. The attribute is for the medication name field only. The value is 1 or 0.

## 1.5 Entries

An entry is a tuple that includes all the seven fields: (name, dose, mode, freq, reason, modifier,

duration). If some fields are not mentioned in the text, their values are “nm”. The exact format of the entry file is not relevant to our annotation task. More detail in Section 5.

## 1.6 Annotation procedure:

The annotation procedure works as follows:

- 1) Open an annotation editor (e.g., Callisto)
- 2) Load the scheme for the task (e.g., choose the \*.jar file that Andy provided)
- 3) Open the file to be annotated
- 4) Annotate the file, “save” frequently (see below)
- 5) When the file is done, “save” it and “export” it.
- 6) Repeat Step 3-5 for other files
- 7) Upload the saved and exported files to Patas under your i2b2 annot/set\*/ directory.

To annotate each field in the text, you need to do the following:

- 1) Mark the span of the field and highlight it.
- 2) Choose one of the seven field tags
- 3) For the “modifier” field, choose the values of three attributes for event/tense/certainty.
- 4) Enter the value for “index”
- 5) If the field is discontinuous, set the “discontinuous” flag to 1
- 6) If the field is a pronoun or “medication”, enter the value for “antecedent”.
- 7) For the “medication name” field, set the value for “output” and “section”.

The first three steps are explained in Section 2, and the last four steps are explained in Section 3. All the attribute values are selected from pull-down menus except for “index” and “antecedent”.

## 1.7 Important guiding principles

Here is a list of principles we shall follow in the guidelines:

- Fields
  - a. Mark all the fields in the input file that you can identify
  - b. A field can be discontinuous; that is, the text is broken into several chunks. In that case, we mark each chunk with the same tag and index and set “discontinuous” flag to 1. Those chunks will be “merged” back when the entry file is produced.
  - c. All the attributes of a field must be specified and the value of “index” can NOT be “nm”.
  - d. A field belongs to n+1 entries, where n is the number of commas in its index. For instance, if its index is “1,2,3”, it belongs to three entries.
- Entry:
  - a. Each entry must have a unique index, and each index must correspond to a unique entry.

- b. The fields in the same entry can appear in different sentences.
  - i. Ex: “He started on drugA last month. He took three pills per day”. The frequency “three pills per day” and the medication name “drugA” are in different sentences, and they belong to the same entry.
- c. All the fields in an entry can be “nm” except for medication name.
  - i. Mark a pronoun or the word “medicine” as “medication name” if and ONLY if it refers to a medication name AND we need to mark it in order to meet the non-empty-medication-name requirement for an entry. More detail in Section 4.6.
- d. The medication name field in an entry should refer to exactly ONE medication.
  - i. Ex1: “He took drugA and drugB”:
 

“drugA” is one field, and “drugB” is another field.
  - ii. Ex2: “He took drugA (another\_name\_for\_drugA)”:
 

mark “drugA (another\_name\_for\_drugA)” as one field.
- e. An entry corresponds to an “event”. There are three kinds of events: “start”, “continue”, and “stop”.
  - i. “change” corresponds to two events: a “stop” event and a “start” event.
  - ii. Ex: In “the dosage of drugA changes from dosage1 to dosage2”, “drugA” appears in two events and therefore its index is “n1,n2”.

## 1.8 The notation in this document

The examples used in the document either are made-up or come from the discharge summaries after simplifying the text.

- drugA/drugB/... in the examples refer to medication names. Similarly, reasonA refers to a reason, and so on.
- Most time we use sgml markups to explain the annotation. For the sake of simplicity, we will include only the attributes that we want to emphasize. For instance, “<n index=“1,2”> drugA </n>” shows only tag and index, not other fields that should be there in the annotated file.

## Section 2: Definition of the seven fields

In this section, we define the seven fields and explain what should and should not be marked. For most fields, we will provide the following:

- Definition
- What is X?
- What is not X?

- Tagging examples
- Open questions

The “what is” and “what is not” lists serve as examples and they are far from complete.

## 2.1 Medication name

**Definition:** They include chemical substances that would help the patient’s condition, e.g., diuretics, as well as biological substances that could help the patient’s condition, e.g., fluids, red blood cells.

### What are medication names?

- Prescribed Medications: Physician ordered, suggested substances.
  - Brand name: e.g., Nexium, Lasix
  - Generics
  - Ingredient name: e.g. furosemide
  - Collective name for a group of medications: e.g., Diuretics, Anthyhipertensives
- Over The Counter Medications: Chemical substances patients can acquire on their own.
  - Brand name: Aspirin
  - Ingredient name: C-vitamin
  - Collective name for a group of OTC medications: Multivitamins
- Biological substances required or suggested by doctors
  - TPN (Total Parenteral Nutrition)
  - Ingredients in TPN if listed are separate medications
  - Blood for transfusion
  - Intravenous (IV) fluid/Saline solution
- Pronouns or the word “medication” which refers to a medication name UNDER certain conditions: see Section 4.6.
- Single medication but two names, when one of the names is parenthetical, results in a single entry:
  - Ex: “TYLENOL (ACETAMINOPHEN) 650 MG PO Q6H PRN Pain”  
Medication name is “TYLENOL (ACETAMINOPHEN)”
- For words such as “nebulizer”, “cream”, “solution”, “tablet” that appear immediately after a medication name, include the words as part of the medication name.
  - Ex: “Lasix tablet 3mg p.o.”  
The medication name is “Lasix tablet”, the mode is “p.o.”
  - Ex: “(Note 71200) DISCHARGE MEDICATIONS: Albuterol nebulizers 250 mg q.4h.”  
The medication name is “Albuterol nebulizers”.
  - Ex: “fluocinonide 0.5% cream applied b.i.d”

“fluocinonide 0.5% cream” is medication name.

What are not medication names?

- Food
- Water
- Diet
- Tobacco
- Alcohol
- Illicit drugs

Tagging Examples:

- “Patient was taking Caltrate plus D”.  
Tag: "Caltrate plus D".
- “Endo: DM II--CZI ss , cont. micronase”.  
Tag: “CZI”
- (Record 166436, lines 27-28) "Override Notice: Override added on 0/26/01 by RONERO ,  
QUINTON I NICHOLAS , M.D. on order for KCL IMMEDIATE REL. PO ( ref # 08069267 )".  
Tag: 'KCL IMMEDIATE REL.' is the drug name.
- "POTENTIALLY SERIOUS INTERACTION: LISINOPRIL & POTASSIUM CHLORIDE  
Reason for override: aware , will check K daily"  
Tag “LISINOPRIL” and “POTASSIUM CHLORIDE” as two medications
- (Record 166436, lines 33-34): "NIFEREX TABLET 50 MG PO BID".  
Tag: "NIFEREX TABLET"
- “ZOLOFT ( SERTRALINE ) 150 MG PO QD”  
Tag: “ZOLOFT ( SERTRALINE )”
- “ECASA (ASPIRIN ENTERIC COATED) 81 MG PO QD”  
Tag “ECASA ( ASPIRIN ENTERIC COATED)”
- “CITALOPRAM HYDROBROMIDE”  
Tag: “CITALOPRAM HYDROBROMIDE”
- “NITROGLYCERIN 1/150 ( 0.4 MG ) 1 TAB SL q5min x 3”  
Tag: “NITROGLYCERIN 1/150”

## 2.2 Dosage

Definition: The amount that a patient takes EACH TIME.

What are dosages?

- 1 TAB
- 0.4 mg
- 1 TAB 04 mg
- One tablet
- 1 bag
- 100 MG
- One unit
- 100 mg x 2 tablets (it means the patient takes 2 tablets each time, each tablet is 100 mg)  
In contrast, in “100mg x2”, “100mg” is dosage, and “x2” is frequency.
- A range of dosage: e.g., 5-10 mg

What are Not dosages?

- “NIFEREX TABLET 50 MG PO BID  
Number of Doses Required ( approximate ): 30”

Tag: “50 MG” as dosage.

Tag “Number of Doses Required ( approximate ): 30” as duration.

Tagging examples:

- “Take drugA 30mg x3 daily”  
“30mg” is dosage. “x3 daily” is frequency.
- If pain does not subside after 3 nitroglycerin, call your doctor or go to the ER.”  
Dosage is “3”.
- “Take upto three sublingual NTG”  
Dosage is “upto 3”
- “NITROGLYCERIN 1/150 ( 0.4 MG ) 1 TAB SL Q5MIN X 3 doses PRN”
  - Dose: “( 0.4 MG ) 1 TAB”
- “Lasix 20mg x1 p.o. q.d”  
“20mg” is dosage.  
“x1” ... “q.d” is frequency. The field is discontinuous.  
“p.o” is mode.

## 2.3 Mode

Definition: Mode or route of administration. Method describing the way the medication is used.

What are modes?

- Oral, orally
- Intravenous
- Topical
- Sulingual
  
- Abbreviation of the above: e.g., p.o. IV

What are NOT modes?

- Words such as “cream”, “tablet”, “solution”, “nebulizer” that immediately follow a medication name, e.g., “Lasix 0.5% cream” together is a medication name.  
Other examples: “Lasix Tablet”, “Saline solution”.

It is true that we can infer the mode from those words. But let’s treat them as part of a medication name if they immediately follow a medication name such as “Lasix”.

Common examples:

- “intravenous (IV) Lasix”

The mode is “intravenous (IV)”

- “Sublingual NTG”

“Sublingual” is mode, “NTG” is medication name

- “The patient was given Lasix both p.o. and IV”

Two modes: one is “p.o.”, and the other is “IV”.

## 2.4 Frequency

Definition: terms, phrases, or abbreviations that describe how often each dose of the medication should be taken.

What is frequency?

- Frequency alone:
  - Ex: “once a day”, “each day”, “daily”, “x3” (it means “three times”), “x3 q.d.”, “b.i.d”
- “as needed”/”prn”/”as necessary” used alone or when it immediately follows or precedes a



frequency: e.g., “three times daily as needed” is a frequency.

Note: if “as needed” follows a duration (e.g., “take drugA for one month as needed”), it is treated as part of “duration”, not as part of “frequency”.

- Temporal phrase that specify when a medication should be taken. These tend to be prepositional phrases. Preposition should be included in the extracted information. One can infer frequency from the temporal phrase.
  - Ex: after meal, after each meal, before bed, at 4pm, with every meal
- Combination of the above:
  - once a day as needed
  - as needed twice
  - as needed after meal
  - three times before meal
  - X3 q.d.
  - X3 daily
  - x3 before meal
  - x3 a day after meal as needed

Tagging Example:

- (file 242336, lines 103-104): "NPH insulin 10 units subcu at 4:30 p.m. pre-dinner"

Tag: f="at 4:30 p.m. pre-dinner"

## 2.5 Reason

Definition: the medical reason for which the medication is stated to be given.

It can be disease, sign or symptom, another medication. Do not tag the text if it is non-medical reason (e.g., financial or insurance considerations).

The reason/indication for a medication can be in a separate sentence than the medication. However, any indication that is not mentioned within a +/-2 sentence window (two sentences before & two sentences after) of the medication should be omitted.

What is Reason?

- Reason can be a symptom: Fever
- Reason can be a disease name: Diabetes
- Reason can be a syndrome (a collection of symptoms or diseases): Acute Lung Injury (ALI)
- Reason can be another medication name: “He was given antihistamines because of his allergic

reaction to penicillin.” Tag: “allergic reaction to penicillin”.

#### What are not reasons?

- Non-medical reasons for which a medication is administered.  
(e.g., Financial or insurance reasons)

#### What “phrase” can a reason be?

- Noun Phrase: Tag the longest base noun phrase as “reason”. The longest base NP has the form “det\* adj\* N+”, it should not include preposition before NP, other PP modifiers, etc.
  - Ex: “He is taking DrugA for his back pain”  
Tag: “his back pain” as reason.
  - Ex: “He was given DrugA for presumed pneumonia”  
Tag “presumed pneumonia” as reason
  - Ex: “His blood sugar was at 185, so we gave 6 units of insulin”.  
Tag “his blood sugar” as reason, not “his blood sugar was at 185”.
  - Ex: “the patient noted that he had a recurrence of this vague chest discomfort as he was sitting and talking to friends. He took a sublingual Nitroglycerin without relief.”  
  
“this vague chest discomfort” is reason.
- Adjective:
  - Ex: “He was given DrugA because he felt dizzy”  
Tag “dizzy” as reason.

#### Tagging Examples:

- (Record 317432, lines 28-29): "apparently in respiratory distress, and treated with oxygen , nitroglycerin , aspirin , Lasix , CPAP , and morphine. Chest x-ray , by".  
  
Tag: “respiratory distress” as reason for each of the six medications.
- (Record 166436, lines 64-66 Lines): “morning of admission when she had onset of left- sided CP which she describes as tightness with radiation to arm while brushing her teeth. Accompanied by racing heart , dyspnea. Somewhat better with SL NTG x 3. Completely alleviated by nitrospray. In ED pain free , no EKG”  
  
There are three reasons: “left-sided CP”, “racing heart”, and “dyspnea”. Treat them as three separate reasons with the same index, not one reason that is discontinuous.
- (Record 317126, lines 8-9): "V6. He had only rare angina since then treated with sublingual nitroglycerin or rest. On 0/4/91 , he had a positive ETT which was"

Tag “rare angina” as reason. Do not include “only” as “only” is an adverb.

- (Record 317126, lines 32-33): "24-48 hours of severe confusion and agitation after extubation requiring four point restraints and Haldol. He gradually resolved"

Tag “severe confusion” and “agitation” as two separate reasons, because “and” is not part of base NP.

One can argue whether “severe” also modifies “agitation”, since either reading is possible, let’s keep the simpler one where “severe” modifies only “confusion”.

- (Record 178031, lines 44-45): "p.o. t.i.d. ( this is day 3 of a 7-day course for reported history of sinusitis ). Patient was discharged to home with followup."

Tag “sinusitis” as reason. A base NP does not include PP, so “history of sinusitis” is not the reason.

## 2.6 Modifier

Definition: Describes (a) what kind event is occurring, (b) when it is occurring, and (c) the certainty of the event.

- The three dimensions correspond to three attributes that exist only for a modifier field (as defined below): event, tense, and certainty.
- The three dimensions are often expressed by different words: e.g., “event” by main verbs, tense by auxiliary verbs, and certainty by words such as “if”, “recommended”.
- To speed up annotation and simplify the procedure, **we will mark only verb group**, which includes the main verb, zero or more auxiliary verbs, and zero or one “to”: e.g., “was taking”, “was recommended to take”.
- If none of three types of events occur, do not mark “modifier” field.

Ex: The sentence “drugA is not working” does not contain a modifier field.

### Tagging examples:

- “He was given drugA”

Tag “was given” as modifier.

He <a index="1" event="start-cont" tense="past" certainty="factual"> was given </a> <n index="1"> drugA </n>

- “He took drugA”

Tag “took” as modifier.

He <a index=”1” event=”start-cont” tense=”past” certainty=”factual”> took </a> <n index=”1”> drugA </n>

- “If you feel sick, take drugA”

<a index=”1” event=”start-cont” tense=”present” certainty=”conditional”>take</a> ....

Tag “take” as modifier. Do not mark “if”.

- “He was recommended to take drugA”

Tag “was recommended to take” as modifier.

<a index=”1” event=”start-cont” tense=”past” certainty=”suggestion”> was recommended to take</a>

- “We switched him from DrugA to DrugB”

Tag “switch” as modifier. You need to tag “switch” **twice**: one for start, one for stop event.

<a index=”1” event=”stop” tense=”past” certainty=”factual”> <a index=”2” event=”start” tense=”past” certainty=”factual”> switched </a> </a> from <n index=”1”> drugA </n> to <n index=”2”> drugB </n>

The three dimensions are explained below.

### 2.6.1 Event: start/stop/continue/start-cont

Definition: Information on whether the medication is started, stopped, or continued. This information is usually expressed in the main verb of the sentence. Mark the event indicated by the main verb or the verb related to the medication.

There are three kinds of events: start, continue, and stop. If it is really hard to determine whether it is a “start” or “continue”, choose “start-cont”. So “event” has four possible values. Event is also expressed as a main verb such as “start”, “continue”, “take”, etc.

Start: start a medication

Example: We started him on DrugA  
He was switched to DrugB.

Common verbs: start, re-start, switch

Stop: stop a medication

Example: We stopped his DrugA.  
We switched him from DrugA to DrugB. (DrugA-m-event: stop,  
DrugB-m-event: start)

Common verbs: stop, switch, “wean slowly”, discontinue

Continue: The medication continues to be administered without any change.

Example: We keep him on DrugA. (DrugA-m-event: continue)

Common verbs: keep, continue

Start-Cont: When we can NOT decide if the medication is new (start) or continue of an old medication (continue).

Example: Being on DrugA helped him.  
Mark “Being” as “start-cont”

Common verbs: be, take, be given

Switch words such as “switch”, “convert”, correspond to two events, and should be marked twice, one with “stop” event and the other with “start” event.

- Ex: “We switched him from DrugA to DrugB.”

Mark “switched” twice, one for “stop”, and the other for “start”.  
The two fields have different indexes.

- Ex: “the patient was started on intravenous Heparin on admission, and was then converted to oral Coumadin.”

Mark “was converted” twice, one for “stop” and the other for “start”.  
The two fields have different indexes.

- Ex: “we decreased his dose of Heparin from 50 mg to 25 mg”

Mark “was decreased” twice, one for “stop” and the other for “start”.  
The two fields have different indexes.

## 2.6.2 Tense: past/present/future

Definition: Information about whether the medication was administered in the past, is being administered currently, or will be administered in the future, to the extent that this information is expressed in the tense of the verbs and auxiliary verbs used to express events. One temporal marker for each event.

Past: The event for the medication is mentioned in past tense.

Example: He took DrugA  
Tag “took” as modifier

Present: The event for the medication is mentioned in the present.

Example: He is taking DrugA.  
Tag “is taking” as modifier

Future: The event for the medication is mentioned in future tense.

Example: He will take DrugA.  
Tag “will take” as modifier.

### **2.6.3 Certainty: conditional/suggestion/factual**

Definition: Information on whether an event occurs. Certainty can be expressed by

- uncertainty words, e.g., “suggested”
- modals, e.g., “should” indicates suggestion.
- Coordination words, e.g., “if”

One certainty for each event.

Conditional: The event for the medication occurs only under certain conditions as mentioned in the text.

Example: If his blood sugar is within 125-150 then give him DrugA in 25 mg dose.  
Tag “give” as modifier. Do not tag “if” for the sake of simplicity.

Suggestion: The event for the medication is/was suggested.

- Ex: “The cardiologist consult suggested DrugA for his arrhythmia”.  
Tag “suggested” as modifier.

Factual: The event is not marked as conditional or suggestion, it is “factual”. This is the default value for “certainty”.

Example: He took drugA. He is given drugA. He will be taking drugA.

## **2.7 Duration:**

Definition: The total length of a medication that is being taken. Duration is often a NP (e.g. 10 days), a PP (e.g., “for one month”) or a clause (e.g., “until the symptom disappears”).

Common Examples:

- An NP:

- Ex: ““Lasix 40 mg p.o. daily x5 days”  
“daily” is “frequency”, “x5 days” is duration
- “use five packs of Lidocaine”  
“five packs” is duration, not dosage, because the patient does not take five packs each time.
- A PP: Please include the preposition
  - Ex: “take drugA for two weeks”.  
“for two weeks” is duration. Include the preposition “for”
- An expression
  - Ex: “take drugA as long as needed”.  
“As long as needed” is duration.
- A clause
  - Ex: “take drugA until the symptom disappear”  
“Until the symptom disappears” is duration
- “As needed” when it follows a duration:
  - Ex: “take drugA for one month as needed”  
the duration field is “for one month as needed”.
- Vague time expressions that indicate the duration of a course of medication such as “temporarily”
  - Ex: “The patient's hospital stay was complicated by frequent PVCs on cardiac monitor , for which he was temporarily treated with Lidocaine”  
  
“temporarily” is duration
- The cumulative dosage:
  - Ex: “5 packs” as in “5 packs of drugA”
  - Ex: “ NIFEREX TABLET 50 MG PO BID.  
Number of Doses Required ( approximate ): 30”  
  
Tag “Number of Doses Required ( approximate ): 30” as “duration”.

### Section 3: Choosing attribute values

As mentioned in Section 1.4, the following are the attributes used in annotation:

- Tag: the type of field
- Subtag: event, tense, certainty (for the modifier field only)
- Index
- Section (for medication name only)
- Discontinuous
- Antecedent
- Output: (for medication name only)

All attributes except index and antecedent have a small set of possible values. We have explained “tag” and “subtag” in Section 2. In this section, we discuss the remaining attributes.

### 3.1. “**Index**”: its value is a positive integer

Indexes are used to link fields together to form an entry; that is, fields with the same index belong to the same entry.

- Each entry has a unique index, and each index corresponds to a unique entry.
- An index is a positive integer. To choose the value of an index, imagine that there is a counter with initial value 0. If the field starts a new entry, increase the counter by 1 and use its value for index. If a field belongs to an existing entry, use the index for that entry.
- If a field belongs to multiple entries, the value of the “index” attribute is the list of indexes joined by commas. For instance, if the field belongs to entry 1 and 2, its index value will be “1,2”.

### 3.2. “**Section**”: its value is “list” or “narrative” (default value)

The attribute “section” specifies whether an entry comes from a list of **medications** or from free text. If it is the former, the value is “list”. If it is the latter, the value is “narrative”.

Note: “list” here means a list of medications, not a list of other entities. For instance, if a medication name appears in a list of actions, it is still marked as “narrative”.

### 3.3. “**Discontinuous**”: its value is 0 (default value) or 1

If a field is not continuous, each chunk of the field will have the same tag, the same index and its value for “discontinuous” is set to 1. For all other cases, the value is 0 (the default value).

Discontinuous fields should be very rare.

### 3.4 “**Antecedent**”: its value is a positive integer or -1 (default value)

Mark a pronoun as medication name only under some condition (see Section 4.6)

If a pronoun is marked as a medication name, the value of the “antecedent” attribute will be the index of the medication name.



Some examples:

- “Take drugA from now on. Take it three times a day.”  
There is only one event. Do not mark "it" as medication name.

- Take drugA from now on. .... Stop taking it after one month.

There are two events. Mark "it" as medication and link it to drugA.

<n index="1">drugA</n> ..... <n index="2" antec="1"> it </n>

### 3.5 “Output”: its value is 0 or 1 (default value)

The attribute indicates if an entry should be included in the entry file. Its default value is “1” indicating the entry will be included in the entry list.

Since an entry must have a medication name, the attribute is shown only in the medication name field.

An entry is included in the final entry list if and only if the text has to include an explicit statement indicating that the patient either took this medication, is taking the medication, is prescribed the medication, is suggested to take the medication, or is otherwise the (past / current / future) experiencer of the medication. Medications taken by other people, and medications that are mentioned but whose experiencer is not known should be marked with output set to 0.

Common examples where the flag is 0:

- Allergy section medications (ALLERGY): The patient lists the medications that he is allergic to.
- The medications taken by someone other than the patient:

Ex: “His father was using DrugA for 30 years for his hypertension.”

- The patient is asked NOT to take a medication.

Note in all these cases, you need to mark the medication field although the field and other fields for the same entry will not be included in the final entry list.

## Section 4: Difficult cases

In this section, we discuss the treatment of some common difficult cases.

### 4.1 The span of a field

The rule of thumb is to keep a field as short as possible as long as the meaning is preserved:

- Medication name: Do not include the determiner (e.g., “the”) before the name

- Ex: “the drugA that he took is effective”

The medication name is “drugA”.

- Duration: Include words such as “for” before the period.
  - Ex: “Take drugA for another three weeks”  
The duration is “for three weeks”
- Modifier: Only mark verb groups as modifiers. A verb group consists of 0 or more auxiliary verbs (e.g., will, be), the main verb, and infinitival “to”: e.g., “will be taking”, “is recommended to take”.
- Reason: Try to keep “reason” field as a longest base NP or an adjective. Mark the whole clause as “reason” if and only if any component of the clause is not sufficient.
  - Ex: “He took drugA for/to alleviate/because of/... back pain”.  
The reason is “back pain”. Do not include the verb or preposition here.
  - Ex: “He took drugA because he has high blood pressure”  
The reason is “high blood pressure”, not the clause “he has high blood pressure”
  - Ex: “He took drugA because his blood pressure is high”  
The reason is “his blood pressure is high”, not “blood pressure”. In this case, you have to choose the whole clause.

## 4.2 Discontinuous field

A field can be discontinuous. In that case, use the same index for the separated chunks.

Common examples:

- Dosage:
  - Ex: “Take two drugA tablets”

“two ... tablets” is dosage.

```
<d index="1" discontinuous="1"> two </d>    <n index="1"> drugA </n>    <d
index="1" discontinuous="1"> tablets </d>
```

- Frequency:
  - Ex: “Lasix 20mg x1 p.o. q.d”

Frequency is “x1 ... q.d”.

- Duration:
  - Ex: “use one pack of Lidocaine until it is finished”

The duration is “one pack ... until it is finished”

### 4.3 The tag of a field

We list some common confusing tag pairs and how to handle them:

#### 4.3.1 “reason” vs. “modifier”

(1) “if S1, take drugA”, where S1 is a clause such as “BS is less than 125”

Mark “take” as “modifier” with certainty=“conditional”. Do not mark part of S1 as “reason”, because we cannot be sure that that is the reason without medical knowledge.

For instance, “if you have a high blood pressure, take drugA”, drugA may or may not be used to treat “high blood pressure”. It could be chosen to treat another disease over an alternative drug which will work for people without high blood pressure.

#### 4.3.2 “dose”, “frequency”, or “duration”

“dose” is how much a person takes EACH TIME, “frequency” is how often a person takes it, and “duration” is how long a person takes it.

Examples:

- “take drugA 20mg x3 daily for one month as needed”  
“20mg” is dose, “x3 daily” is frequency, “for one month as needed” is duration
- “As needed” is tricky, as it can be part of a frequency or part of a duration.

So the rule of thumb is that if it follows a duration (e.g., “for one month”), treat it as part of duration. In all other cases, treat it as part of frequency.

### 4.4 One entry or multiple entries?

The “index” attribute can be seen as the “index” of an entry in our internal entry file. So there is one index per entry. Therefore, choosing the value of “index” is the same as determining the number of entries. Here are some common cases.

#### 4.4.1 When there are multiple drug names

- 1) When the drug names refer to different drugs, the drug names should have different indexes.
  - Ex: “take drugA and drugB, 1 tablet each daily”

There are two entries: one for drugA, and one for drugB.

... <n index="1"> drugA</n> and <n index="2"> drugB </n>, <d index="1,2"> 1 tablet  
</d> each <f index="1,2"> daily</f>

- 2) When the drug names refer to the same drug, there is only one medication field with one index.
- Ex: “take dragA (another\_name\_for\_drugA), 1 tablet daily”

```
... <n index="1"> drugA (another_name_for_drugA) </n>, <d index="1"> 1
tablet
</d> <f index="1"> daily </f>
```

#### 4.4.2 When there are multiple reasons

- 1) When they are the reasons for taking the same drug, the reasons should have the same index.
- Ex: “He is taking drugA for reason1 and reason2”

Reason1 and reason2 have the same index.

```
<n index="1"> drugA </n> .... <r index="1"> reason1 </r> and <r index="1"> reason2
</r>
```

- 2) When they are the reasons for taking multiple drugs, the reasons should have multiple indexes, one for each drug. This is consistent with the decision for 2) in Section 4.4.1.

- Ex: “He is taking drugA and drugB for reason1 and reason2”

Both reasons have two indexes as there are two drugs.

```
<n index="1"> drugA </n> and <n index="2"> drugB </n> for <r index="1,2">
reason1 </r> and <r index="1,2"> reason2 </r>
```

#### 4.4.3 When there are changes in medication name, mode, dosage, frequency, etc., there should be multiple entries.

- 1) Change of medication names

- Ex: “We switched him from drugA to drugB”

“switched” is the modifier for both entries. You need to mark “switch” twice as its event values are different.

```
<a index="1" event="stop" tense="past" certainty="factual"> <a index="2" event="start"
tense="past" certainty="factual"> switched </a> </a> him from
<n index="1"> drugA</n> to <n index="2"> drugB</n>
```

- 2) Change of mode

- Ex: “We switched him from oral Lasix to IV”

Once again, “switched” needs to be marked twice.

<m index="1"> oral </m>     <n index="1,2"> Lasix </n>     to   <m index="2"> IV  
</m>

- Ex: "The patient was given Lasix both p.o. and IV"

Two entries: one with dosage "p.o", and the other with dosage "IV".

### 3) Change of dosage

- Ex: "We increased his DrugA from 25 mg to 50 mg"

DrugA appears in two entries.

< n index="1,2"> DrugA </n> from     <d index=1> 25 mg </d>     to <d index=2> 50 mg  
</d>

- "Insulin on sliding scale. Take 6 units if BS > 185. Take 4 units if BS > 140."

There are two entries, one with "6 units", and the other with "4 units" as dosage.

### 4) Change of modifier

- Ex: "Lasix 40 mg p.o. daily for five days and then discontinue"

There are two entries, one with "take-continue" event, the other with "stop" event. "Lasix" appears in both entries with different indexes.

## 4.5 Handing "switch" words

"Switch" words often signal two events: one stop event and one start event.

See the examples in Section 4.4.3.

## 4.6 Marking pronouns and the word "medicine/drug" as medication names

Let's look at an example. Suppose the text is "He took drugA last month. ... He will continue taking it for a few more months". There are two events here: one for each sentence. For the second event, we could mark "drugA" as its medication name, but if we do that, the medication name field could be far away from other fields for the second event.

To avoid this problem, we mark "it" as medication name, give it its own index, but set its antecedent value to be the index given to drugA. In general, we mark a pronoun or a word such as "medicine" as medication name only in this scenario; that is, there are two entries: one of them uses a medication name, and the other uses a pronoun or the word "medicine" to refer to it.

Some examples:

- Ex: “He has been taking drugA since last month. He takes it three times per day.”

There is only one event. Do not mark “it” as medication name.

- Ex: “He took drugA last month. He will continue taking it for a few more months.”

There are two events, one for each sentence. Mark “it” and coindex it with “drugA”.

`<n index="1"> drugA </n> ..... <n index="2" antecedent="1"> it </n>`

- Ex: “He is taking drugA. The medication is very effective.”

There is only one event. The second sentence does not correspond to a start/continue/stop event. So do not mark “medication” as medication name.

Note: A single medication referred to with different names on different occasions requires separate entries for each mention. Do not try to resolve NP co-reference or synonymy.

- Ex: “The patient was started on **TYLENOL ( ACETAMINOPHEN )** 650 MG PO Q6H PRN Pain. Acetaminophen was to be used for five days... Tylenol was stopped.”
  - Medication for entry 1: “**TYLENOL ( ACETAMINOPHEN )**”
  - Medication for entry 2: “Acetaminophen”
  - Medication for entry 3: “Tylenol”

## 4.7 Coordination

Coordination is notorious hard, and very often there are no easy tests to determine what the right answer is. You should use your best judgment. If multiple readings are possible, choose the “simplest” one, as explained below.

### 4.7.1 One field or two fields?

In a coordination structure “A and B”, you have to determine whether the string is one field or two fields: For instance, in “he took drugA for severe confusion and agitation”, is “severe confusion and agitation” one reason or two reasons? The rule of thumb is that if there is no convincing evidence to show that “A and B” together is one concept, treat it as two. In this case, we will treat it as two reasons, and therefore two fields. This is also consistent with our requirement that “reason” should a base NP most of the time.

If we decide there are two fields, the next question is whether “severe” modifies only the first field or both. Once again, if both readings are possible, choose the one that does not require a discontinuous field; that is, let’s assume that “severe” only modifies the first field, not both.

In summary, in this example, there are two reasons: “severe confusion” and “agitation”.

More examples:

- Ex: “Take drugA and drugB for reason1 and reason2”

“reason1” and “reason2” are two reasons for both drugA and drugB. Therefore, each reason has two indexes.

```
<n id="1">drugA</n> and <n id="2">drugB</n> ....  
<r id="1,2"> reason1</r> ... <r id="1,2"> reason2</r>
```

- Ex: “Take drugA for reason1 and drugB for reason2”

“reason1” is for “drugA” only, and “reason2” is for “drugB” only, so each reason has one index only.

```
< n id="1"> drugA</n>   for <r id="1"> reason1</r> and  
<n id="2">drugB</n>   for <r id="2"> reason2</r>   ...
```

## Section 5: Generating entry files from annotated files

This section provides some background information. You don’t need to read it for your annotation task.

Our internal entry file is formed from the annotated file in two steps:

- Extract the fields from an annotated file: `cat annotated_file | extract_field.sh > field_file`
- Combine fields to form entries: `cat field_file | form_entry.sh > entry_file`

### 5.1. The field file

The format of field file is:

```
start_pos end_pos tag="tag" index="nn" section="xx" discontinuous="xx" ... text="..."
```

Each line in the field\_file corresponds to a marked chunk in the annotated file; that is, each line corresponds to a field in the text with two exceptions:

- If a chunk appears in the same field but the field belongs to multiple entries (i.e., its index is a list of integers), the chunk corresponds to only one line in the field file.

Ex: “Take drugA and drugB daily”

```
Start_pos end_pos   tag="f"   index="1,2"   .... text="daily"
```

- If a chunk is marked multiple times as it appears in multiple fields (the fields have different indexes and different values for some attributes such as “event”), it corresponds to multiple lines in the field file: one line for each field.

Ex: The annotated text is “<a index=“1” event=“stop” tense=“past” certainty=“factual”> <a index=“2” event=“start” tense=“past” certainty=“factual”> switched </a> </a>” as in “he switched from drugA to drugB”

The chunk “switched” corresponds to two lines in the field\_file:

```
start_pos end_pos tag="a" index="1" event="stop" tense="past" certainty="factual" ....
text="switched"
```

```
start_pos end_pos tag="a" index="2" event="start" tense="past" certainty="factual" ....
text="switched"
```

- When a field is discontinuous, each chunk in the field corresponds to one line in the field file.

Ex: “Take two drugA tablets daily”

```
Start_pos end_pos tag="d" index="1" discontinuous="1" ... text="two"
Start_pos end_pos tag="d" index="1" discontinuous="1" ... text="tablets"
```

## 5.2 Our internal entry file

Each line in the entry file corresponds to an index in the annotated file. The entry file is formed by merging lines in the field file which have the same index.

The format of the entry file is:

```
index="idx" || n="medication name" {x} || d="dosage" {x} || f="freq" {x} || m="mode" {x} || r="reason"
{x} || a="modifier" {x} || dur="duration" {x}
```

- Ex: index="1" || n="bicarb" 11:1 11:1 || d="two amps" 11:2 11:3 || f=nm || m=nm ||  
r=nm || a=nm || dur="three months" 11:5 11:6
- idx is an integer.
- If a field is present in the text, the following {x} is "start\_pos end\_pos", where start\_pos and end\_pos are the position of the field.
- If a field is not present, {x} is not there and the field value is nm without quotation marks.
- If there are multiple reasons with the same index, the line will have multiple r="reason":
  - Ex: r="reason1" start\_pos end\_pos || r="reason2" ...
- If a field is discontinuous, the separated chunks are concatenated in the entry file.

## 5.3 The final entry file



The format of the final entry file depends on the requirements from i2b2. Here are some possible differences between our internal entry file and final entry file:

- Possible differences in formatting
- The final entry file might use character offset, not word offset.
- When an entry has multiple reasons, the entry could correspond to multiple lines in the final entry file.
- When a reason is a clause, the i2b2 guidelines might either treat it as nm or choose an NP in the clause as the reason.
- The treatment of “it” vs. “medication” for co-reference.
- In the final entry file, do not include offset for modifier.

Our annotation should be rich enough so that we can automatically create the final entry file from the internal entry file.