SAM Patents – Annotation Guidelines

Thomas Heitz, Yaoyong Li, Kalina Bontcheva

March 24, 2009

1 Introduction

This document defines guidelines for annotating patents with metadata, such as references to figures and other patents. As the whole process is one of annotation, the resulting data can be classified as meta-data relating to the original documents (i.e. the patents). In this document, data and meta-data are used interchangeably to refer to the semantic meta-data that enriches the patents.

The annotation itself should be carried out with the GATE TEAMWARE Annotation Editor. To launch this, first login at `http://ldc.matrixware.com:7233/ldc-safe/executive/` and click on the "Open Annotator Editor" link. Once open, click on the first button to get an annotation task. Further training on how to use this UI will be provided to the annotators prior to them attempting new annotation tasks. General help with the annotation GUI can be found by clicking on the *Information icon* in the toolbar at the top.

In addition, all GATE Teamware help documentation is available at `http://ldc.matrixware.com:7233/ldc-safe/executive/helpInfo.html`.

2 General Guidelines

The format of this document consists of a section per type of annotation and a list item per annotation. There may also be a *Specific Guidelines* subsection at the end of some section to detail some difficult cases to annotate. The format of each list item is:
• **Annotation**: description of the annotation.
Example:
text of the example `[annotated expression]` *text of the example*
Note that the annotation text in the example is enclosed by the square brackets "[" and "]".

Annotate the complete span for each annotation. For example, for a reference to patent or literature, please always include all the information such as author, title and publication date into the annotation. Some examples are:
as disclosed in `[Japanese Unexamined Patent Publications of Tokkai Shou 54-119377 and Tokkai Shou 55-54198]`
e.g., in `[Tokiyuki Yoshida, Shinichi Shindo, Tadayoshi Ohgaki and Kiyoshi Nakayama compiled, Kaimen Kasseizai Handbook ( Handbook of Surfactants ), Kogaku Tosho Shuppan Co., Ltd. (1987)]`.

Conjunctive phrases mentioning two or more objects of the same annotation type should be tagged as one annotation, including the conjunction and punctuation, if those references share the same header. For example, *Figure 1 and 2*; *Figures 1-3*; *Figures 1 to 10* should all be annotated as one Reference annotation of type Figure. For more details about how to annotate conjunctive phrases, please see the explanations about multiple references in the beginning of Section 4.

For some data types, there might be pre-annotated references in documents. These need to be checked and corrected if needed.

If you are unsure and want to bring something to the attention of the curator, set the *requires-attention* field to true on the respective annotations. You can moreover add a free text comment in the *comment* field.

3 Document Sections

Document section annotations segment patent documents into sections in a flat and consecutive manner. A patent document has three main parts, *the first page containing* the bibliographic data and abstract , *the descriptions part*, and *the claims part*. The description part usually contains several sub-sections, which should be annotated separately (details below).

At the very top of the document is its identifier, which is not a Section and should not be annotated. Example: US20010001827-A1-20010524-2118. This is the only part of the document, which does not need to be covered by Section annotations.

Annotate all the consecutive paragraphs that belong to a section as a whole Section annotation. The first paragraph does not always contain a title. For each Section annotation added automatically, their validity should be checked carefully and corrected, if necessary.

As Section annotations are of different types, one needs to select the correct type from the list of alternatives below.

If a paragraph is not already covered by a Section annotation, then a new one must be created and given one of the types below or the type Other if unsure.

1. First Page
   (a) **BibliographicData**: the information at the start of the patent document which contains bibliographic data. It is usually followed by the section *Abstract*, so if in doubt, determine the beginning of the abstract, then annotate everything before the abstract as `BibliographicData`.
   (b) **Abstract**: a short text disclosing the gist of the solution of the problem through the invention, and the principle uses of the invention. The abstract should be differentiated from the summary of the invention section, which appears lower down in the patent's text, if at all.
   (c) **Sponsorship**: this optional section would discuss the relationship between the patent and any funding received from funding bodies for the scientific research. An example title is:
   STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT.
   (d) **CrossReferenceToRelatedApplications**: this optional section contains references to other related patents. If present at all, it would have a title such as:
   RELATED APPLICATIONS, CROSS-REFERENCE TO RELATED APPLICATIONS or INCORPORATION BY REFERENCE....
2. Description of the invention
   (a) **BackgroundArt**: a section describing the background art of the invention. It is often the first sub-section of the description part. It may have a title like:
   BACKGROUND OF THE INVENTION or *Background Art*
   or begin with a sentence like:
   The present invention relates in general to ....
   However, if a title BACKGROUND OF THE INVENTION is followed immediately by another title e.g. *Field of the Invention*, then the field sub-section should be annotated as `TechnicalField`.

i. **TechnicalField**: specifies the technical field to which the invention relates. It may have a title like:
TECHNICAL FIELD or FIELD OF THE INVENTION.

ii. **PriorArt**: the existing body of technological information against which the current invention is judged to determine if it is patentable as being novel and not obvious. It may have a title such as:
PRIOR ART.

(b) **SummaryOfInvention**: gives a summary of the invention, e.g. the objects and advantages of the invention. It may have a title as:
SUMMARY OF THE INVENTION or BRIEF DESCRIPTION OF THE INVENTION.
It should be distinguished from the patent's abstract, which appears at the beginning of the document.

(c) **DisclosureOfInvention**: this section describes the invention in such terms as the technical problem, the technical solution, and advantageous effects if any. It may have a title like:
OBJECT OF THE INVENTION.

(d) **DrawingDescription**: this section describes briefly the drawings (if any). It may have a title like:
Brief Description of the Drawings or DESCRIPTION OF THE DRAWINGS or LEGEND.

(e) **DetailedDescription**: This section typically has a title such as:
DETAILED DESCRIPTION OF THE INVENTION or DETAILED DESCRIPTION.

(f) **TechnicalProblem**: the technical problem which the invention tries to solve. It may not be an independent section in some patents, and instead may just be indicated by some sentences, in which case it does not need to be annotated as an independent section.

(g) **TechnicalSolution**: the solution adopted by the invention to deal with the technical problem. It may not be an independent section in some patents, and instead may just be discussed in a few sentences, in which case it does not need to be annotated as an independent section.

(h) **Effects** annotation: the effect of the invention. It would be a stand-alone section in many Japanese (JPO) patents, but may not be an independent section in most others. There is no need to annotate it independently if that's the case.

3. Usage of the invention

(a) **BestMode**: this section defines the best means known to the applicant of practising the invention claimed. This should be done in terms of examples, where appropriate, and with reference to the drawings, if any. Patents from some countries may not contain this section. Where available, it may have a title such as:
Best Mode.

(b) **PreferredEmbodiment**: This section typically has a title such as:
DESCRIPTION OF THE PREFERRED EMBODIMENTS

(c) **Examples**: this section contains examples of how to use the patent and typically has a title such as:
EXAMPLE X (with X a number).

(d) **UsageOfInvention**: this section indicates explicitly the way in which the invention is to be exploited in industry and the way in which it can be made and/or used. The section may be absent in some patents if the usage is obvious from the description or nature of the invention and typically has a title such as:
INDUSTRIAL UTILIZABILITY.

4. Claims of the invention

- **Claims**: this is typically the last section in the patent document and defines the claims for which protection is sought. It is one of the three main parts of patent documents and is always present (albeit sometimes without an explicit section title). Typical titles are CLAIM or CLAIMS or it may begin with sentences such as:
  What is claimed is: or *What is Claimed:* or *I claim:*.

5. Bibliography

   - **Bibliography**: this section contains literature or patent references used in the document. It typically has a title such as:
     BIBLIOGRAPHY or REFERENCES or PUBLICATIONS

6. Unknown section

   - **Other**: this section type should be used when the annotator cannot find any type which is appropriate or they are unsure of their choice, as these will be checked later by a domain specialist.

Specific Guidelines:

- Most patents only contain some of the section types listed above. At the very least there are sections for the bibliographic data, the abstract, the patent description and the claims at the end.
- If a section contains one or more sub-sections, e.g. the `BackgroundArt` section contains the two sections `TechnicalField` and `PriorArt`, then only annotate the more-specific, embedded sections.
- Most sections have section titles which are useful when determining the type for the Section annotation. However, if a patent document does not have section titles, then one has to read the text in order to segment it into sections.
- The sub-sections in the description part of each patent can be grouped according to their contents. For example, `BackgroundArt`, `TechnicalField` and `PriorArt` are about the background of the invention; `DisclosureOfInvention`, `TechnicalProblem`, `TechnicalSolution`. `DetailedDescription` and `DrawingDescription` are about the invention itself; and `BestMode`, `PreferredEmbodiment` and `UsageOfInvention` are about the usage of the invention.

Quick Reference for Section Titles:

Section title	**Section type**
ADVANTAGES OF THE INVENTION	*Effect*
Background Art	*BackgroundArt*
background of the invention	*BackgroundArt*
BENEFITS OF THE INVENTION	*Effect*
Best Mode	*BestMode*
brief description of the invention	*SummaryOfInvention*
CITED LITERATURE	*Bibliography*
cross-reference to related applications	*CrossReferenceToRelatedApplications*
description of the drawings	*DrawingDescription*
description of the preferred embodiments	*PreferredEmbodiment*
detailed description	*DetailedDescription*
detailed description of the invention	*DetailedDescription*
example 3	*Examples*
EXEMPLARY PROCEDURE	*Examples*
EXPERIMENTAL PROTOCOLS	*DetailedDescription*

field of the invention	*TechnicalField*
I claim:	*Claims*
incorporation by reference...	*CrossReferenceToRelatedApplications*
INDUSTRIAL UTILIZABILITY	*UsageOfInvention*
legend	*DrawingDescription*
object of the invention	*DisclosureOfInvention*
OTHER PUBLICATIONS	*Bibliography*
PREVIOUS TECHNOLOGY LEVEL	*PriorArt*
prior art	*PriorArt*
REFERENCE MATERIALS	*Bibliography*
related applications	*CrossReferenceToRelatedApplications*
statement regarding federally sponsored research or development	*Sponsorship*
summary of the invention	*SummaryOfInvention*
technical field	*TechnicalField*
What is Claimed:	*Claims*
What is claimed is:	*Claims*

Example Sectioning

Some section annotations are shown in the screenshots in Figure 1.

4 References

Reference annotations are used for parts of text that refer to either objects in the current document (e.g. figures, tables, etc.) or to other documents (other patents, scientific papers, etc.).

A reference annotation consists of two parts, a header indicating the type of the refence, and one or more ids of reference which may be in number or letter or a combination of the two. For example, in *Figure 1 and 2* the header is *Figure* and the ids are *1* and *2*. In *U.S. Pat. No.3,765,999* the header is *U.S. Pat.* and id is *No.3,765,999*.

If Reference annotations have an id set, leave it as it is. If not present, do not fill it in.

Conjunctive phrases mentioning references to two or more objects of the same type should be tagged as one Reference annotation, including the conjunction and punctuation. For example, *Figures 1 and 2*; *Claims 1-3*; *Tables 1 to 10* should all be annotated as one Reference annotation of type Figure, Claim or Table.

One feature called `multiple` is used for distinguishing multiple reference from single reference in one Reference annotation. It has three values, "false" referring to one single reference such as *claim 3*; "enumeration" listing all reference ids such as *Fig. 1 and 2* and *Examples 2, 4 and 5*; "interval" showing an interval of ids sharing the same header such as *Figures 1-5* and *Tables 3 to 6*.

There are several types of Reference annotations and one must always select the correct one from the following list according to the header of the reference:

- **Figure**: refers to figures in the patent document. Examples:
  `[Figure 2]` *schematically illustrates a side view of a lifting aid of the "tile lift" type.*
  the enlarged view of `[FIG. 7]` *in addition to* `[FIG. 4]`.
  as a segment of the full-length ACT-4-h-1 sequence shown in `[FIG. 5]`.
  portions of the loop of Henle (`[Fig. 2E]`*)*

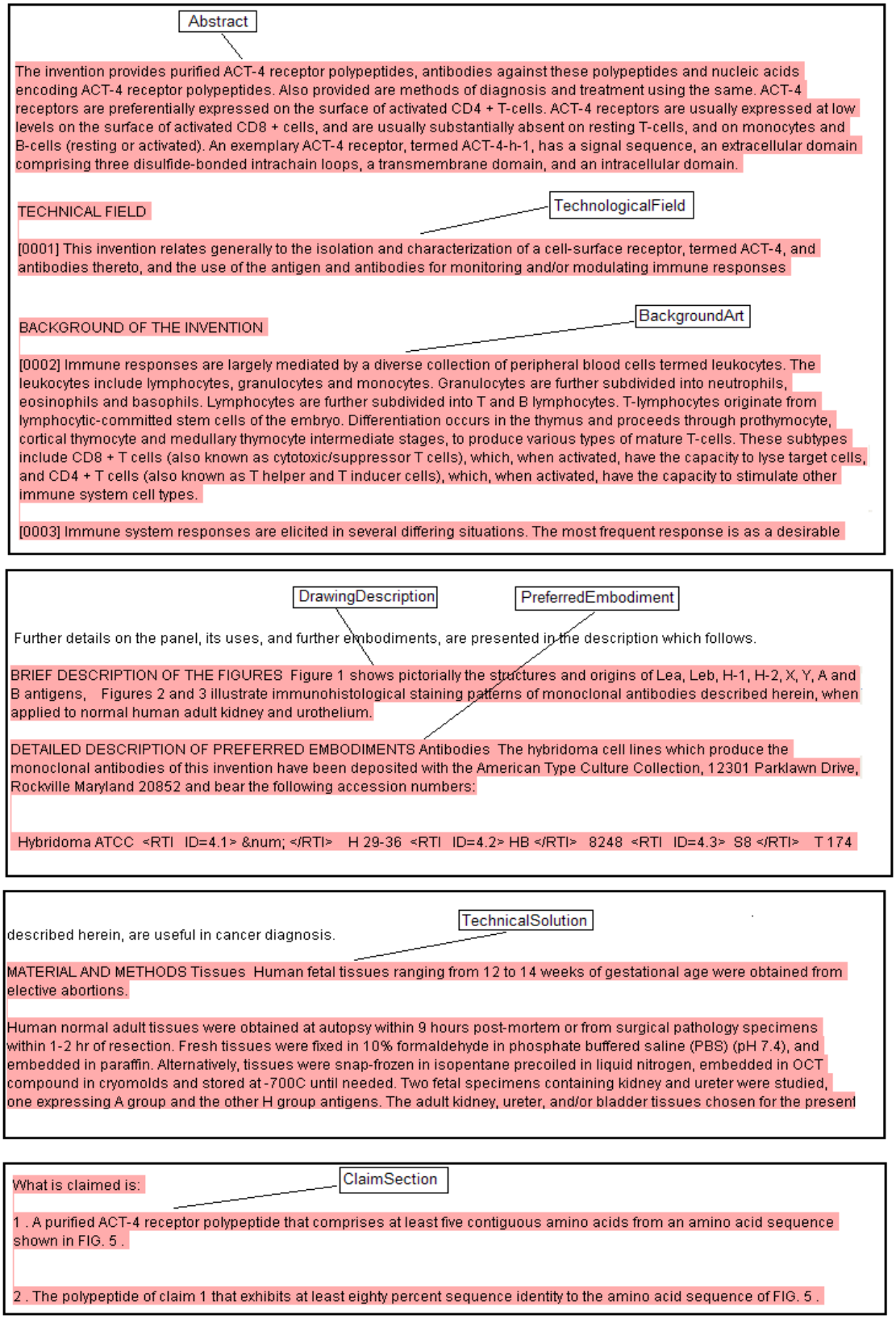
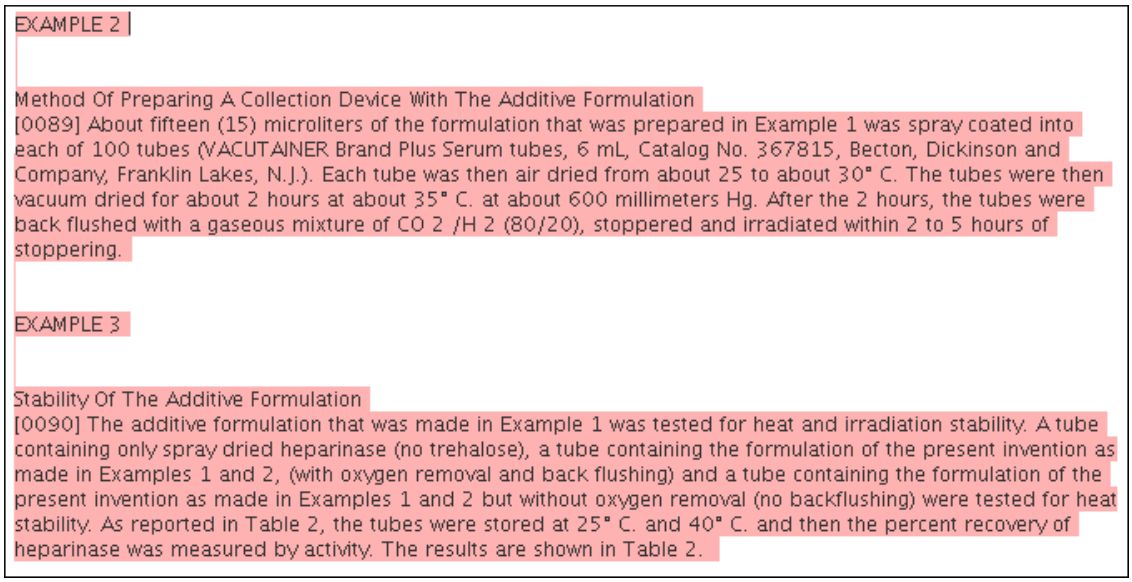


Figure 1: Four pieces of text with some instances of sections.


EXAMPLE 2

Method Of Preparing A Collection Device With The Additive Formulation
[0089] About fifteen (15) microliters of the formulation that was prepared in Example 1 was spray coated into each of 100 tubes (VACUTAINER Brand Plus Serum tubes, 6 mL, Catalog No. 367815, Becton, Dickinson and Company, Franklin Lakes, N.J.). Each tube was then air dried from about 25 to about 30° C. The tubes were then vacuum dried for about 2 hours at about 35° C. at about 600 millimeters Hg. After the 2 hours, the tubes were back flushed with a gaseous mixture of CO 2 /H 2 (80/20), stoppered and irradiated within 2 to 5 hours of stoppering.

EXAMPLE 3

Stability Of The Additive Formulation
[0090] The additive formulation that was made in Example 1 was tested for heat and irradiation stability. A tube containing only spray dried heparinase (no trehalose), a tube containing the formulation of the present invention as made in Examples 1 and 2, (with oxygen removal and back flushing) and a tube containing the formulation of the present invention as made in Examples 1 and 2 but without oxygen removal (no backflushing) were tested for heat stability. As reported in Table 2, the tubes were stored at 25° C. and 40° C. and then the percent recovery of heparinase was measured by activity. The results are shown in Table 2.

Figure 2: Examples section.

- **Table**: refers to the table in document. Examples:
  in the second transcriptional cassette (`[Table 1]`*)*
  although there were some significant differences (`[Table II]`*).*

- **Equation**: refers to the mathematical equation or formula in document. Examples:
  reference marks shown in the `[expressions (1) and (2)]`

- **Claim**: refers to claim made in document. Examples:
  Use according to `[claim 1]` *characterized in that said vaccines*
  A transfer according to any one of `[claims 1 - 6 and 9]`

- **Formula**: refers to the chemical formula in document. Examples:
  in connection with `[formula (I)]`
  Catalysts having the `[formulas (I) and (II)]`

- **Patent**: refers to the external patents cited in document. Examples:
  copending `[U.S. Patent Application Serial No. 474,415, "Monoclonal antibody S8 is known to detect B-antigen," Ueda, et al., PNAS 78:5122 (1981)]`.
  Such dyes are described in `[U.S. Patent 4,524,128 (Edwards et al.)]`,
  `[JP 2001-109101 (Adachi)]`
  The `[international patent application WO 92/07990]` *discloses a possible use of a colour copier*
  disclosed in `[U.S. Pat. No. 5,429,211]`,
  is disclosed in `[U.S. Pat. No. 5,487,768 to Zytka et al]`.
  `[U.S. Pat. Nos. 5,713,791 and 5,344,365]`
  (`[Hattersley et al ., Cartilage Induction by Bone Morphogenetic Proteins , U.S. Patent No. 5,902,785]`*).*

  In some document, there might be references to patents at the beginning of the documents, in the bibliographic section. Please do not annotate them, because the preprocessing can obtain those annotations from the original markups reliably. See the Specific Guidelines below for more details.

- **Literature**: refers to the external literatures (e.g. published article and book, and unpublished manuscript) cited by the patent. Examples:
  `[Rettig, et al., Cancer Res. 45:815 (1985)]`
  `[Sakamoto, et al., (unpublished manuscript)]`
  `[Hudson & Hay, Practical Immunology (Blackwell Scientific Publications, Oxford, UK, 1980), Chapter 8]`
  `([Kawaguchi et al. J. Biol. Chem., 264: 5762-5767, 1989])`

  Note that there are several company names in the following text such as "Keeling & Walker Ltd.", which should not be annotated as literature references.

  "Preferred electrically conductive particles are conductive Sb-doped tin oxide powders commercially available from a number of sources (e.g., Keeling & Walker Ltd., Dupont Ishihara Sangyo Kaisha, Performance Chemicals, Mitsubishi Metals, Nissan Chemical Industries, etc.)."

  A sequence of literature references, such as a bibliography or references list at the end of the patent document, or simply enumerated references, should be annotated all together as 1 Reference annotation, with *type* Literature and *multiple* feature with value *enumeration*. For example, all this is 1 annotation:

  `[D.L. Stemple and N.K. Mahanthappa, Neural stem cells are blasting off, Neuron 18:1-4 [1997]; Y. Renoncourt et al., Neurons derived in vitro from ES cells express homeoproteins characteristic of motoneurons and interneurons, Mechanisms of Development 79:185-97 [1998]; A.J. Kalyani et al., Spinal cord neuronal precursors generate multiple neuronal phenotypes in culture, J. Neurosci. 18(19):7856-68 [1998].]`

- **Example**: refers to an example section. Examples:
  the start codon for the polypeptide (see `[Example 1]`*) and/or a polyA sequence*
  protease inhibitory assays such as, e.g. `[Example 2]` *or Heidtmann et al., 1990, Clin Chem 36: 2077-2081.*

Specific Guidelines:

- **Bibliographic section at the beginning**: please note that you do **not** need to annotate any references in the bibliographic data section which is normally at the beginning of document. If you are not so sure which is the bibliographic data section, select the annotation type Section and the bibliographic data section is covered by the Section annotation with the "type" feature as "BibliographicData".
- **Enumeration of references**: if several references of the same type appear in a conjunction or a list, they should be annotated as one annotation. Don't forget to set the multiple feature of the annotation as the one of the three values "false", "interval" and "enumeration", as explained in the begining of this sub-section.
- **Patent reference with title, author, etc.**: when annotating patent references, please not only include the information about country of origin, patent number(s) and dates but also, if present, the inventor name or title of the patent itself. In general, references to other patents usually tend to give at least a patent number, which sometimes also contains a date (e.g. *1223445/19871201*). Sometimes the country of origin is also included (e.g., US, JP, GB). In some cases the authors and title of patent are also included. Some examples are:

  in `[U.S. Patent 4,524,128 Edwards et al]`.
  in `[Korean laid open utility model application No. 1999-007692]`,
  as disclosed in `[Japanese Unexamined Patent Publications of Tokkai Shou 54-119377 and Tokkai Shou 55-54198]`
  (`[Hattersley et al ., Cartilage Induction by Bone Morphogenetic Proteins, U.S. Patent No. 5,902,785]`*).*

- **Foreign patents**: references to German Published Application XXX or "German Utility Model" should be annotated as references to other patents, because in Germany and Austria these are patent-like documents[1].

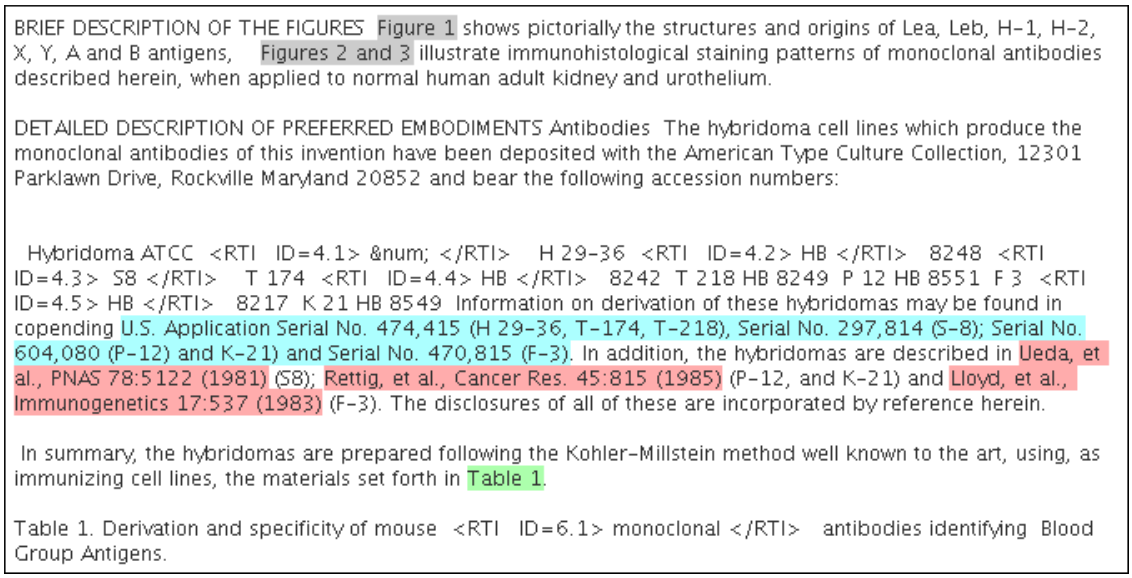
- **Title and caption for table, figure, etc.** : even if a reference occurs inside the caption or the title of a table, figure, etc. it should be annotated as a `Reference` of type `Table`, `Figure`, etc.

- **Bracket around references**: literature and Patent references sometimes have brackets around them or being as part of them. If the entire reference is inside a pair of brackets, then don't include the brackets in the annotation. However, if part of the reference is out of the brackets and another part is inside in the brackets, then the annotation should include the brackets. See the following examples:

  see `[J.Smith (ACME Journal of Research (2003))]`
  (see `[J.Smith (ACME Journal of Research (2003))]`*)*

Detailed Example:

See figure 3.


BRIEF DESCRIPTION OF THE FIGURES Figure 1 shows pictorially the structures and origins of Lea, Leb, H-1, H-2, X, Y, A and B antigens, Figures 2 and 3 illustrate immunohistological staining patterns of monoclonal antibodies described herein, when applied to normal human adult kidney and urothelium.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS Antibodies The hybridoma cell lines which produce the monoclonal antibodies of this invention have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville Maryland 20852 and bear the following accession numbers:

Hybridoma ATCC <RTI ID=4.1> # </RTI> H 29-36 <RTI ID=4.2> HB </RTI> 8248 <RTI ID=4.3> S8 </RTI> T 174 <RTI ID=4.4> HB </RTI> 8242 T 218 HB 8249 P 12 HB 8551 F 3 <RTI ID=4.5> HB </RTI> 8217 K 21 HB 8549 Information on derivation of these hybridomas may be found in copending U.S. Application Serial No. 474,415 (H 29-36, T-174, T-218), Serial No. 297,814 (S-8); Serial No. 604,080 (P-12) and K-21) and Serial No. 470,815 (F-3). In addition, the hybridomas are described in Ueda, et al., PNAS 78:5122 (1981) (S8); Rettig, et al., Cancer Res. 45:815 (1985) (P-12, and K-21) and Lloyd, et al., Immunogenetics 17:537 (1983) (F-3). The disclosures of all of these are incorporated by reference herein.

In summary, the hybridomas are prepared following the Kohler-Millstein method well known to the art, using, as immunizing cell lines, the materials set forth in Table 1.

Table 1. Derivation and specificity of mouse <RTI ID=6.1> monoclonal </RTI> antibodies identifying Blood Group Antigens.

Figure 3: Reference annotations. From top to bottom: Figure, Figure, Patent, Literature, Literature, Literature, Table.

5 Measurements

All measurements should be annotated as Measurement annotations, which have a type, which needs to be assigned as one of the following elements in the list.

Furthermore, unit measurements have a `nunit` property that need to be set to the normalized unit for this unit. For example *%*, *percent*, etc. must have a `nunit` set to `ratio`, *hour*, *minute*, etc. must have a `nunit` set to `second` and so on.

[1] http://en.wikipedia.org/wiki/Gebrauchsmuster

- **scalarValue**: scalar measurements are the most typical type of measurements and consist of a numeric value that can be a number word. They should **always** be related to a measurement unit such as inches or mm, which must be annotated as a separate measurement annotation of type `unit`. See figure 4 and these examples:
  - *Sections were treated for* `[30]` *minutes in* `[1]` *% hydrogen peroxide*
  - *plasticizing point of about* `[160]` *C., a hardness of -* `[93]` *°Shore and an elasticity of* `[400]` *% is particularly useful*
  - *The focal adhesion contacts vary in size and distribution, but* `[80]` *% of them measure less than* `[6]` *um* `[2]` - Note: to read as 6.2 micrometre.
  - *The deposition chamber is evacuated to a pressure less than or equal to* `[2x10 -7]` *Torr.*
  - *the local surface profile of central lumen 130 average less than about* `[0.0015]` *inch and preferably less than about* `[0.0005]` *inch* - Note: There is no interval here.
  - *using a disposable, plastic insulin syringe and* `[29G×½]` *" needle*
  - `[Two]` *to* `[four]` *weeks later, spleen cells were harvested and analyzed*

- **discreteValue**: discrete measurements are the types of measurements where the value is one of a given set of possible values. Their is no measurement unit related to them. Examples:
  - *A small shirt size is* `[S]` *in USA.*
  - *This hotel is rated* `[**]`.

- **unit**: a unit that should **always** be related to at least one scalar measurement, which is annotated separately as a `scalarValue`. Do not include a final dot in the unit. See figure 4 and these examples:
  - *Sections were treated for 30* `[minutes]` *in 1*`[%]` *hydrogen peroxide*
  - *Deposition rates up to 20* `[nm/sec]` *are achievable using ion beam-assisted*
  - *a fatigue life of 400 MM* `[cycles]` *as measured by simulated accelerated testing,* - Note: to read as 400 millions (double M for plural) cycles of a certain time constant.

- **interval**: an interval measurement contain two `scalarValue` and zero, one or two `unit` or two `discreteValue` without `unit`. Do not include any word before the first value and after the last unit. An interval can also contain two other intervals. Beware that '± 2200 bp' or 'SD ± 54.8' without any adjacent value are not intervals but ranges referring to another value. See figure 5 and these examples:
  - `[100-200 g]` *sheared heterologous DNA or tRNA,* `[0-10%]` *dextran sulfate,* `[110 5 to 110 7 cpm/ml]` *of denatured probe ...*
  - *a temperature between about* `[300 and 1100 degrees Centigrade]`
  - *a hyperthermal energy between* `[0.1 eV and about 700 eV]`*, preferably between* `[5-50 eV]`
  - *a plasticizing point in the range* `[120-160 C]`*. containing dispersed fine particles.*

- **compound**: compound measurements are measurements that comprise a group of other measurements. A good rule of thumb is that all measurements covered by compound measurements should have the same dimension (time, length, etc.) but different names of unit. Examples:
  - *Transcription was evaluated at the site of injection by semi quantitative RT-PCR analysis, after* `[2 days and 1 week]`.

Specific Guidelines:

- **List of values**: should be annotated separately as `scalarValue` or `discreteValue`. Example:
  - `[3]`, `[5]`, `[6]` *cm.*
  - *for* `[15]` *seconds, and* `[20]` *cycles of* `[15]` *seconds for each step of denaturation*
  - *The turbulence intensity may be greater than* `[0.055]`, `[0.06]`, `[0.07]` *or up to* `[0.10]` *or* `[0.20]` *or even greater.* - Note: 'turbulence intensity' should be annotated as `unit`.

- **Relative unit**: Percentage (%), ppm, ratio (:, /), etc. should be annotated as `unit` and the value as `scalarValue`. Examples:

- *wherein the NF3/CHF3 ratio is approximately* `[9:7]` - with '9' and '7' as `scalarValue` and ':' as `unit`.
- *93 °Shore and an elasticity of* `[400%]` *is particularly useful* - with '400' as `scalarValue` and % as `unit`.
- *A NF3/CHF3/N2 gas mixture with a* `[11/8.6/80.4]` *ratio in percent* - with '11', '8.6' and '80.4' as `scalarValue` and the two '/' between those number as `unit`. 'ration in percent' should also be annotated as a `unit`.

- **Repetition of unit**: if a unit is repeated in another form like first the long form then the short form in bracket, annotate only the closest to the value. Example:
  - *23* `[millimetres]` *(mm).*

- **Repetition of value**: if a value is repeated in another form like first in word then in number, annotate only the closest to the unit. Example:
  - *About fifteen (*`[15]`*) microliters of the formulation.*

- **Interval with approximative units**: words like 'about', 'approximatively', 'circa', 'around' can be used in front of measurement inside intervals. Include only the second one of those words. Examples:
  - *an angle of approximately* `[30 to 60]`
  - *between* `[0.1 eV and about 700 eV]`
  - *The formulation provides substantial recovery of about* `[50% to about 60%]` *post irradiation..*

- **Things that are not preprocessed**: compound measurements, discrete value, unit and value that are not adjacent, interval of ratios, ratio between measurements, units without SI counterpart like "bp" for base pair in genetics or "parts by weight" are not always preprocessed but you should annotate them in case we want to add them to the unit database.

Concrete Examples:

[0030] A special variation of the glue layer comprises laminating a thermoplastic moulded polyurethane sheet on the transparent or white elastomer layer. A 100 µm thirk transparent or coloured aromatic polyester film having a plasticizing point of about 160° C. a hardness of 93° Shore and an elasticity of 400% is particularly useful for the purpose. The silk screen printed polyurethane layers and the polyurethane sheet can be laminated together at 160° C. under slight pressure, so that the sheet does not melt, but only adheres to the applied layer. During application of the finished transfer to a textile, which takes place at 200° C. and 320 kPa in 12 seconds, the polyurethane sheet melts and forms a very strong glue layer between the textile and printed image.

Figure 4: Excerpt of a patent document with couples of measurement annotation where the first element is a scalar measurement and the second element is the unit of the measurement.

Di-t-butyl peroxide in a concentration range of 0.1% to 10%, Di-cumylperoxide in a concentration range of 0.1% to 10%, t. butyl cumyl peroxide in a concentration range of 0.1% to 10%, all percentages are calculated by weight related to dry weight of fibres, the step wherein compression and decompression is carried out in an range of 5-2-5-2-5 atms., in a range of time of 1 to 15 minutes, temperature in a range of 20°C-120°C, said wood or vegetable fibres in a concentration range of 1% to 30%, the step wherein complementary chemical addition is carried out in a concentration of 1% - 5% of sulfonating or oxidating agents at the entry in to steam pressurized vessel, temperature in a range of 130°C to 220°C, saturated steam pressure in a range of 0.5 MPa to 2.5 MPa, time in a range of 10 seconds to 30 minutes, wherein the step of defibration is carried out in decompression chamber to atmospheric pressure, the step wherein diffusion washing or chemical addition of sulfonating, bleaching or complexing agents are followed by the step wherein atmospheric refining is carried out in a concentration range of 5% to 25%, freeness in a range from 700 to 50 CSF (Canadian Standard Freeness), or the step wherein pressurized refining is carried out in a pressure range of 1.5 - 15 atms., in a consistency range of 5% - 40%, freeness in a range from 700 to 50 CSF followed by the step of non compulsory washing, screening and cleaning of the produced pulp, the step wherein bleaching of refined pulp is carried out employing one or more of the bleching agents:

Figure 5: Excerpt of a patent document with interval measurement annotations. The expression *10 seconds to 30 minutes* is also a compound measurement.