Guidelines for the Annotation of General Scientific Concepts

by
Maria Liakata and Larisa Soldatova

Department of Computer Science
Aberystwyth University, UK

15 December, 2008

Table of Contents

1 Introduction

These guidelines discuss the annotation of General Scientific Concepts (GSCs) in the context of the ART project (http://www.aber.ac.uk/compsci/Research/bio/art/). The goal of the ART project is to provide an authoring tool to automatically translate scientific papers from a free text format into a semantic format where the GSCs are designated in the text [Soldatova *et. al.*, 2007]. A batch of manually annotated papers will serve as training data for the development of the tool, which aims to be of use to authors, editors and reviewers of scientific papers. Authors will be able to use it to port their paper to a semantically rich machine readable format, which can be easily queried by natural language processing tools and other applications. At the same time it will help them check the completeness of their argumentation by establishing the presence and distribution of GSCs in their papers. The ART tool can also be used to guide authors and editors in structuring the papers and help reviewers by providing a concise representation of the content.

We focus on how to recognise and annotate GSCs in papers presenting experiment driven scientific research. Our main assumption is that a scientific paper represents a scientific investigation and therefore needs to contain the key concepts for the description of scientific investigations. These GSCs constitute a subset of the EXPO ontology [Soldatova & King, 2006]. The latter is a resource developed at Aberystwyth University to represent scientific experiments in a theoretically sound manner consistent with the requirements of the Semantic Web. The usefulness of the GSCs and their adequacy in describing scientific papers have been evaluated in an on-line survey, in the context of the ART project (http://www.aber.ac.uk/compsci/Research/bio/art/news/).

In the following we present the GSCs and give detailed guidelines for marking them manually in text. It is very important that annotators read the guidelines carefully and adhere to them during annotation, as precision and consistency are paramount to obtaining clean, re-usable data with high inter-annotator agreement. We will use *italic font* for the example's text. GSCs and their properties will be placed in angled brackets further on in the text.

2 General Scientific Concepts

GSCs are a set of concepts essential for the semantic representation of scientific investigations. These concepts were chosen after a preliminary annotation of papers and were evaluated by researchers during an online survey. They consist of the following:

1	<Conclusion>
2	<Result>
3	<Goal>
4	<Method>
5	<Object>
6	<Experiment>
7	<Observation>
8	<Hypothesis>
9	<Motivation>
10	<Background>
11	<Model>

The first 7 concepts constitute the **core GSCs**, representing the minimum information required for the semantic representation of the content of a paper [Soldatova & Liakata, 2007]. GSCs can have properties, and if present, these properties have values. For example, <Method> of investigation has the property <Novelty> with values <New> (if the method is introduced in the current paper as a new method) or <Old> (if there is mention of the method having previously been used by other researchers). <Method> can also have the property <Advantage> with values <Advantage> and <Disadvantage>. Knowledge of property values is crucial to the full comprehension of the investigation. Therefore, for the accurate representation of a scientific investigation we require both the specification of the GSCs and the values of some of their essential properties, namely:

<New>, <Old>, <Advantage>, <Disadvantage>

GSC concepts stem from an ontology of science [EXPO], where they have been defined as possessing more properties. However, only the two properties mentioned previously and the above four values are considered for the purposes of the current work. According to the definition of GSCs, only <Method> and <Object> of the investigation can have the above properties. Note that <Object> does not have the property <Old> as the object of an investigation is by definition <New>.

Users are required to assign an identifier (concept ID) to each new GSC found in the text, and this concept ID can be assigned to several different sentences, referring to the same GSC. Section 4 provides a detailed explanation of each GSC and gives examples of how to annotate them.

3 Annotation Procedure

3.1 Before Annotation

It is important to read the entire paper (provided in .pdf format) before the annotation as the meaning of some sentences only becomes apparent in the context of the whole argumentation of the paper. The aim is not so much to understand the solutions/methods proposed in detail but to get a general feel of the paper and argumentation structure and focus on the main points:

- what issues the paper addresses;
- what methods are used/proposed;
- what the main results/conclusions are.

It is crucial to read the entire paper to obtain this information as the abstract is not always well-written and does not constitute an accurate summary of the paper.

3.2 During Annotation

When annotating, one should bear the following in mind:

- Each sentence should be annotated with a single GSC label.

- A **paper title** is not necessarily a sentence but we allow its annotation in the same way. It often includes the hypothesis of a paper (see Section 5.5), a goal or conclusion.
  However, it is not always meaningful, in which case it should be annotated as <Background>.

- The same Identifier (concept ID) should be used to link sentences related to the same GSC. This applies to associating e.g. <Advantage> of a method with the <Method> or e.g. two sentences referring to the same method. More instructions for the assignment of identifiers (concept IDs) is provided in the annotation algorithm, Section 3.3.

- The same type of GSC can appear in different sections of the paper (e.g. the <Object> can be mentioned in the Abstract, Introduction and Conclusion sections). It is important to mark a GSC **everywhere** one encounters it. Annotators are asked to mark what is in the text and whether it can be assigned a GSC label, not to do selective information extraction.

- You may judge that different parts of the sentence correspond to different GSC concepts. Example :

  "*The direct measurement of gaseous HONO with the LOPAP-analyser clearly shows that a large fraction of NO2 taken up was converted into HONO, in line with the experiments shown above and with an earlier study on dark reactions of NO2 on anthrarobin [27]*".

  Here we have both an <Observation>: "The direct measurement of gaseous HONO with the LOPAP-analyser clearly shows that a large fraction of NO2 taken up was converted into HONO," and a <Result>, which is the comparison of the current evidence with previous work "in line with the experiments shown above and with an earlier study on dark reactions of NO2 on anthrarobin [27]". In such **cases choose the concept with the highest priority and assign that to the sentence**. The significance order of GSC labels is represented in the table in Section 6. Thus, the sentence above would be annotated as <Result>.

- A paper doesn't have to include all GSCs. It may be the case e.g. that there is no explicit <Hypothesis> but there still can be instances of <Experiment>, <Model>, <Method>, <Conclusion> etc. However, each paper should definitely include **at least one instance** (but more likely several) of the following GSCs:

  <Conclusion>
  <Result>
  <Object> (or <Goal>)
  <Method> (normally followed by advantages and disadvantages).

  No paper annotation can be complete without the above concepts. If one has failed to detect them, one should process the paper again.

3.3 Annotation Algorithm

Annotation is sentence based. This means that annotators are asked to mark each sentence in the text with the appropriate GSC concept and assign the corresponding concept ID and properties (if applicable). The annotation tool, SAPIENT[1] should be used to aid the annotation process. SAPIENT will split the paper text into sentences and display them so that the structure of a paper on a section level is maintained. However, individual paragraphs will no longer be distinguishable.

[1] SAPIENT stands for "Semantic Annotation of Papers: Interface & ENrichment Tool"

Therefore, we recommend that during the annotation of a paper annotators have a **printout** of it handy or that the pdf version of the full paper is accessible on a computer screen.

To assure consistency between annotators and facilitate the process, the following annotation algorithm should be followed:

Consult the paper.pdf or printout and proceed on a paragraph basis, since by reading each sentence in isolation it is difficult to assess whether it forms part of a GSC. Here there is the implicit assumption that a paragraph constitutes a conceptually coherent discourse unit. Follow these steps:

1. Read a paragraph. Then work on each individual sentence of the paragraph.
2. Proceed to see if you can answer any of the questions (a)-(f) below about the current sentence, using the tree diagram (see fig.1).
3. When you arrive at a leaf node question querying the assignment of a GSC label to the sentence do the following:

    3.1. If there is a GSC label you can attach, do so. Check the sentence prior to the current one; assign the same concept ID if:

        i. the GSC of the current sentence provides a continuation of the concept in the previous sentence;

        ii. the GSC of the current sentence is a property (<Old>/<New>; <Advantage>/<Disadvantage>) of the GSC assigned to the previous sentence with a GSC label.

    3.2. Consider previous sentences and sections. If you are **absolutely sure** that there is a sentence which is not in the immediate vicinity of the current sentence (e.g. is in a different section of the paper) but refers to the same concept as the current one, then assign to the current sentence the concept ID of the earlier co-referring sentence.

    3.3. If neither 3.1 nor 3.2 apply or you are not sure, **give a new concept ID** to the GSC. In either case continue to the next sentence in the paper and step (2).

    3.4. At the end of the current paragraph go back to step (1) to proceed with the next paragraph.

    3.5. If you experience difficulties assigning a single label to a sentence, look at the priority given to GSC labels in Section 6.

3.4 Tree Diagram of Questions

Questions (a)-(f):

(a) Does this sentence contain information about the current work or pre-existing work ?
If the answer is 'current work', continue to question (b);
If the answer is 'pre-existing work', continue to question (c);

(b) Does this sentence talk about the objectives of the current work, the actual approach, the outcome or neither?
In the first case, go to question (d);
In the second case, go to question (e);
In the third case, go to question (f);
In the last case assign the label <Background>

(c) Does this sentence talk about an existing method?
In thise case consider assigning the GSC tag: <Method> - <Old>. Follow the process (steps 1-3) listed above. You can specify if the sentence is more specifically about the advantages or disadvantages of the method. See more details in Section 5.7.
If the sentence does not refer to a pre-existing <Method> then label it as <Background>.

(d) Does this sentence talk about a goal, motivation, object of the investigation or hypothesis?
In each of these cases consider assigning the relevant GSC tag: <Goal>, <Motivation>, <Object> or <Hypothesis>. Follow the process (steps 1-3) listed above. In the case of an <Object> you can specify if the sentence is more specifically about the advantages or disadvantages of the object.

(e) Does this sentence talk about the theoretical framework/model in this approach, a new method or an experiment?
In each of these cases consider assigning the relevant GSC tag: <Model>, <Method> - <New>, or <Experiment>. Follow the process (steps 1-3) listed above. In the case of a <Method> you can specify if the sentence is more specifically about the advantages or disadvantages of the method.

(f) Does this sentence talk about an observation, a result or a conclusion of the current approach?
In each of these cases consider assigning the relevant GSC tag: <Observation>, <Result>, or <Conclusion>. Follow the process (steps 1-3) listed above.

Figure 1 shows a diagram of the questions for the annotation (a) – (f). More details about questions (a) – (f) are provided in the following section. More information about how to label sentences with GSCs are provided with the annotation interface SAPIENT.

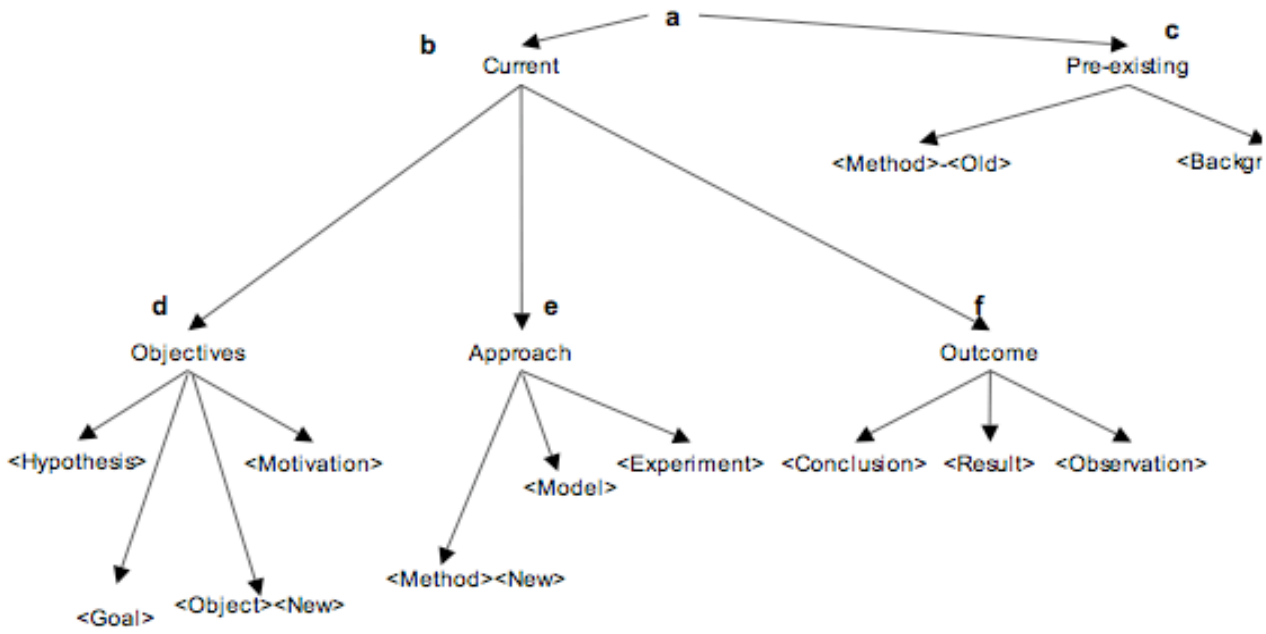

Figure 1: *A question diagram for the annotation procedure.*

4 Details about the questions

(a) Does this sentence contain information about the current work, pre-existing work or neither?

Typically, when authors refer to pre-existing work they use expressions such as *'widely used', 'widely accepted', 'in recent...'*. Also neutral statements about a fact or problem can point to pre-existing work. The existence of citations usually constitutes a telling sign of reference to pre-existing work. As a rule, pre-existing work will be mentioned in the Introduction and Background sections of the paper. However, references to methods already used by other researchers or by the authors can appear anywhere in the paper apart from the conclusions section.

Current work should make up a large part of the paper and is usually presented in the first person plural, along with a great amount of detail with respect to methods employed, results and conclusions drawn. Previous work produced by the same authors can also be mentioned in first person plural but it still constitutes pre-existing work. The use of present tense and expressions containing *'new'* can be clues pointing to current work.

(b) Does this sentence talk about the objectives of the current work, the actual approach or the outcome?

Once you have established that the sentence you are looking at is mentioning current work, you need to decide whether it fits under one of the above three categories. The objectives include: any assumptions the investigation is seeking to prove (<Hypothesis>), any tasks to accomplish(<Goal>), any specific entities about which the investigation seeks to discover knowledge (<Object> of investigation) and why the investigation is important (<Motivation>). If you ascertain that the sentence is about objectives, go to question (d).

The actual approach includes: any theoretical model(s) employed to aid with the investigation of the objectives (<Model>), the method employed (as well as its advantages and disadvantages) <Method><New>, any specific experiments <Experiment>. If you ascertain that the sentence is about objectives, go to question (e).

Finally, the outcome of the investigation can be expressed as direct observations from an experiment <Observation>, as the interpretation of observations <Result> or more general statements stemming from the findings of the investigation <Conclusion>. If you can tell that the sentence is about the outcome of an investigation, go to question (f).

(c) Does this sentence talk about a background fact or an existing method?

The assumption here is that you know that the sentence in question refers to previous work related to the current investigation but which of the previous two cases does it fall under?

<Method> - <Old>

An existing method is defined in the same way as a new method, in the section where we deal with question (e) (see below). The difference is that an existing method is more likely to be found in the literature review section and to be accompanied by citations. If it is not obvious that the method used is a new method it is safer to assume that it is an existing method.
[For details see ***Section 5.7***]

<Background>

<Background> is the label to use for any background information (about pre-existing, already known entities) which is important for the current investigation.
[For details see ***Section 5.1***]

(d) Does this sentence talk about the goal, motivation, object of the investigation or hypothesis?

The assumption here is that you know that the sentence in question refers to the objectives of the current investigation. But which one of the above four is it?

<Goal>

Sentences which talk about what an investigation aims to show, what problem it aims to solve or explore should be annotated with the tag <Goal>. A goal of an investigation is the target state of the investigation where intended discoveries are made, approaches are tested, problems are demonstrated, tasks formulated, etc.
[For details see ***Section 5.4***]

<Object>

An <Object> of investigation is the principal entity of an investigation, - what this investigation is about. An <Object> of investigation is an entity or series of entities about which the authors seek to gain new knowledge by means of the investigation. There can be several objects of investigation in a paper. An object of investigation often appears in the Abstract or the Introduction section of a paper as well as the Conclusion section. However, it is not uncommon for objects of investigation to be mentioned in the Results and Discussion sections as well. An object of investigation is often introduced as the object of verbs such as '*report', 'discuss', 'investigate', 'study'* or as subjects of these verbs in passive voice. An object of investigation can also be signalled by an *'is a'* expression, providing a definition of the object.
[For details see ***Section 5.10***]

<Motivation>

A sentence that indicates why the investigation is important should be annotated as <Motivation>. Motivation for an investigation is the stimulus for achieving the goal of the investigation, the reason to carry out the investigation.
[For details see ***Section 5.9***]

<Hypothesis>

A hypothesis can refer to an investigation, experiment or model. A hypothesis is rarely stated explicitly in text but can usually be inferred from the goals and objects of an investigation. If there is mention of what authors aim to prove in their investigation or experiment, then mark the sentence containing it with the GSC <Hypothesis>.
[For details see ***Section 5.5***]

(e) Does this sentence talk about the model in this approach, a new method or an experiment?

The assumption here is that you know that the sentence in question refers to the actual approach of the current investigation. But which one of the above four concepts is it?

<Model>

A sentence which describes a theoretical model used or formed within an investigation as well as how the model is represented should be annotated with the label <Model>. A model will be normally described in the Methods or Results and Discussions section. It can be expressed as a system of equations, a set of principles, the design of an object, or a set of assumptions.
[For details see ***Section 5.8***]

<Experiment>

If a paper reports an investigation which includes 'physical' experiments involving instruments and measurements, then one should assign the label <Experiment> to the sentences which provide the direct description of these experiments. The direct description of experimental design, experimental factors, protocols, materials and equipment should all be annotated as <Experiment>. Sentences which provide a generic description of experiments should be labelled as <Method>. Sentences which mention experiments which do not form part of the investigation reported in the currect paper, should be annotated with the label <Metod> - <Old>.
[For details see ***Section 5.3***]

<Method> - <New>

A method of an investigation is the means by which the authors seek to achieve a goal of the investigation. We consider an experimental method as a special case of <Method> due to its importance to scientific investigations. All sentences which provide a direct description of experiments within the context of the current investigation should be marked as <Experiment>. There are three cases when the tag <Method> should be used:

- Sentences which describe methods, other than experimental methods should be annotated as <Method>. Sentences which provide a non-direct, generic description of experimental methods should be annotated as <Method>.
- Sentences which mention or describe experiments which do not constitute part of the investigation reported in the current paper should be annotated as <Method>.

[For details see ***Section 5.6***]

(f) Does this sentence talk about an observation, a result or a conclusion of the current approach?

The assumption here is that you know that the sentence in question refers to the outcomes of the current investigation, which will also normally occupy the largest part of a scientific research paper. Observations, results and conclusions will appear all throughout the paper apart from the Introduction and Background sections. The three of them are related through a type of generalisation relation (observation < result < conclusion). Sometimes the boundaries between them may be hard to draw. The following will help clarify the distinction between them. Note that some examples for the annotation of these three types of outputs are joined together.

<Observation>

An <Observation> encompasses the data/phenomena recorded within an investigation, as well as how the data are represented. An experimental observation is a direct observation of nature or a set of values, prior to interpretation. For example, measurements or direct descriptions of the outcomes of an experiment or method qualify as observations. Observations are often signalled by the verbs *'show', 'observe', 'see', 'detect'* and the inclusion of quantities, data, figures and measurements as well as their numeric or non-numeric description.

[For details see ***Section 5.11*** ***Section 5.12*** ***Section 5.2***]

<Result>

A <Result> of the investigation is a set of findings, obtained through the interpretation of the observations and available evidence .

[For details see ***Section 5.12*** ***Section 5.11*** ***Section 5.2***]

<Conclusion>

Conclusions are statements inferred from observations and results, to support or reject a research hypothesis. A conclusion should answer the question of whether the target or targets of an investigation have been achieved, whether a hypothesis has been confirmed or refuted and what new contribution to knowledge has been obtained through the investigation.
[For details see ***Section 5.2*** ***Section 5.11*** ***Section 5.12***]

5 The GSC Concepts in detail

5.1 <Background>

<Background> is the label to use for any background information (about pre-existing, already known entities). For example, label a sentence as <Background> if it contains background related to the current work, usually just before or after an annotated <Method>.

Please use this category sparingly and after you have made sure that a sentence does not fall into any of the other categories.

Example 1:

> *"The positions of the absorption bands measured for [Cu(4-picoline)n]2+ complexes appear slightly blue shifted with respect to those recorded for pyridine over the same size range. In a related study of $Ni(X\text{-}py)_4(ClO_4)_2$ complexes and with a range of substituent X on pyridine, Bull and Moore concluded that the effects of substitution at the 3- and 4-positions were too small to influence the magnitude of the ligand field splitting constant, Δ[41]."*

The 1st sentence "*The positions (..) range*" is an <Observation> of the current investigation. The following sentence gives background information pertaining to the particular observation about pyridine.

Example 2:

> *"Although Ag(II) complexes are frequently unstable in solution, they can often be stabilised with nitrogen-containing heterocyclic organic ligands such as pyridine and pyridine derivatives.[22,42,43]. These complexes frequently exhibit four coordination and examples, such as [Ag(py)4]S2O8, are reported to be square planar.[22,42] In the condensed phase, λmax is ~470 nm for [Ag(py)4]S2O8,[22] which is to be compared with ~600 nm for a typical Cu(ii)/pyridine complex[1]. For the clusters in the range of n = 2–7, photofragmentation patterns were investigated over the wavelength range 450–1100 nm."*

The 3rd sentence "*For the clusters (..) nm*" is an <Object> of investigation, since it charachteristically mentions that *"photofragmentation patterns were investigated"*. The previous three sentences all give background information about Ag(II) pyridine complexes, whose photofragmentation patterns are investigated. These three sentences should each be marked as <Background> and should be given the same concept ID.

Example 3:

> *"We first investigated the replacement of OH groups in the hydroxyapatite structure by an oxygen ion and water molecule, in effect the products of the dissocation of one hydroxy ion and the addition of the resulting proton to the second hydroxy ion. This reaction is a necessary intermediate step in the dehydration process, before removal of the volatile water from the channel[23,24]."*

The 1st sentence "*We first (..) hydroxy ion.*" should be labeled as <Object> of investigation as it talks about a reaction being investigated. The sentence "*This reaction (..) [23,24*]" provides background information about the reaction which is the object of investigation and should therefore be annotated as <Background> with a new concept ID.

5.2 <Conclusion>

Conclusions are statements inferred from observations and results, which support or reject a research hypothesis or summarise the findings of an investigation. A conclusion should answer the question of whether the target or targets of the investigation have been met, whether a hypothesis has been confirmed or refuted and what new contribution to knowledge has been achieved through the investigation. Conclusions are not direct interpretations of observations in the way that results are but rather present the broader picture and the overall impact of the investigation.

There are many fewer instances of <Conclusion> in an investigation than there are instances of <Observation> or <Result>. <Conclusion> is at the highest level of abstraction from results and observations and represents the distilled knowledge or outcome stemming from the intermediate findings of the investigation. Instances of <Conclusion> are mainly encountered in the Abstract and Conclusion but also in the Results and Discussion sections.

Example 1:

> *"The effect on the SERRS spectrum when a SERRS Beacon and a complementary sequence where allowed to hybridise is shown in Fig. 4. The resulting SERRS spectra were still dominated by FAM but a broad fluorescence background was also observed. This indicated that the FAM was removed from the silver surface. This initial study clearly demonstrates the high potential that SERRS Beacons have as a new class of molecular reporter."*

In this example one can see the build up from <Observation> to <Result> and then to <Conclusion>. The authors first look at the actual spectra produced, then interprete it and then finally arrive to new knowledge about the object that produced the spectra, namely the SERRS Beacon. The first and the second sentences should be labelled as <Observation> with the same concept ID, since the second sentence continues to describe the same set of observations. The third sentence *"This indicated (..) surface*" should be assigned the label <Result>, because it is an interpretation of the observations. Finally, the last sentence is a <Conclusion> since it infers new knowledge by summarising and generalising over the results and observations.

Example 2:

> *The results from these experiments (Table 2) show that the phototransformation of NO2 -> HONO is also effective when the electron donor is not a phenolic compound. This indicates that neither radical oxidation intermediates of the phenols nor products of the phenolic oxidation (for example quinones) are necessary for the photoreduction of NO_2 -> HONO. This is a strong indication that the diphenylketyl radical (protonated and/or deprotonated form) is a key species in the reduction of NO_2. Furthermore, the experiment with iodide showed that energy transfer reaction between triplet 4-BBA and the proble phenols in the ground state (eqn. (9)) is not essential in the reaction system.*

The first sentence, "*The results (..) compound*" interprets the experimental results (the observations) and should therefore be annotated as <Result>; there is a clear indication for this choice in the text: "*the results…show"*. The second sentence, "*This indicases (..) HONO*" is a <Conclusion> drawn from the results. The following sentence, "*This is (..) NO_2"* is also a <Conclusion> originating from the same results, but it is a different one so it should be assigned a different concept ID to that of the previous sentence. The last sentence talks about the results of an experiment and should be annotated as <Result>. This example shows that even though there is usually a build-up from observations to conclusions, it is not unheard of to have results and conclusions interspersed in the text.

Example 3:

> *"Based on the results reported a mechanism is suggested, in which photosensitised electron transfer is occurring.".*

This sentence constitutes a <Conclusion>. It is a high-level interpretation of experimental observations, giving a mechanism-model as a new contribution to knowledge.

More examples of <Conclusion> can be found in the following:

Example 4:

"Within the three types of metal catalysts used—Au, Ag and In, Au was found to provide the best control for achieving rod-growth even though the melting point of bulk gold is significantly higher then the reaction temperature."

Example 5:

"Room and low temperature absorption and photoluminescence measurements show that the band-gap shifts to the red upon increasing rod length revealing strong quantum confinement along the long axis in InAs rods, providing spectral coverage of the near-IR range relevant for telecommunication applications."

Note that the distinguishing factor between a <Result> and a <Conclusion> in this example is the potential for applications stemming from the findings as well as the generalisation over all experimental conditions.

Example 6:

"This proposes that during the growth stages, gold nanoparticles melt or dissolve thereby being removed from the rods. Such Au then may aggregate to form the large Au particles detected in the percipitate."

Example 7:

"Therefore, we may assume that the general contribution of non-radiative processes is similar. The nature of this process, and detailed spectroscopic investigation of the optical properties of InAs rods is currently in progress."

Example 8:

"Therefore, we propose that NO2 is reduced in the present system by reductants produced during the photosensitized oxidation of phenols."

Example 9:

"This indicates that neither radical oxidation intermediates of the phenols nor products of the phenolic oxidation (for example quinones) are necessary for the photoreduction of NO2→ HONO. This is a strong indication that the diphenylketyl radical (protonated and/or deprotonated form) is a key species in the reduction of NO2."

Example 10:

"Our calculations of calcium-deficient hydroxyapatites show that charge-compensation by carbonate groups is energetically favourable, where the calcium-deficiency can either occur as vacancies or as monovalent ions substituting at the calcium positions."

5.3 <Experiment>

If a paper reports an investigation which includes 'physical' experiments, involving instruments and measurements, then one should assign the label <Experiment> to the sentences which provide the direct description of these experiments. The description of experimental design, experimental factors, protocols, materials and equipment should all be annotated as <Experiment>. In some cases a whole section will be assigned the label <Experiment> but is useful to keep track how many different experiments are described so as to assign different concept IDs to them where appropriate.

An experiment usually is described as a process with explicit mention of conditions, requirements, instruments, experiment steps, etc. The label <Experiment> is reserved only for those sentences which describe experiments that are part of the current investigation. The description of experiments that are not part of the current investigation should be annotated as <Method>. Sentences talking about other types of methods (analytical, theoretical) should be annotated as <Method>. All generic, non-direct, descriptions of experiments should also be annotated as <Method> .

Experiments are usually described in detail in the Methods section of the paper. However, mentions of experiment can also be found in the Abstract and Results and Discussion sections.

Example 1:

> *"In this collaborative study, the uptake kinetics and product formation rate were measured by different flow tube reactors in combination with a sensitive HONO instrument."*

This sentence talks about the equipment used for obtaining measurements from an experiment conducted within the context of an investigation ("*study*") . It should therefore be annotated as <Experiment>.

Example 2:

> *"Using the combination of a Pyrex flow tube and mercury lamp, we are initiating photochemistry at wavelengths longer than 300nm. Further experiments were performed with specific and sensitive HONO detection, at low concentrations of NO2 ".*

Here both sentences should be labelled as <Experiment> and they should both have the same concept ID as they are talking about the same series of experiments with the same instruments and configuration.

Example 3:

> *"As an example a complementary sequence with overhanging bases to mimic hybridisation to a target sequence within a longer strand of DNA was added to the SERRS Beacon and conditions for optimal SERRS examined".*

This sentence should be labelled as <Experiment> since it describes an experimental procedure including conditions and inputs.

5.4 <Goal>

Sentences which talk about what an investigation aims to show, what problems it aims to solve or explore, should be annotated with the label <Goal>. A goal of an investigation is the target state of the investigation where intended discoveries are made, approaches are tested, problems are demonstrated, tasks formulated, etc.

An investigation can have several different goals and a goal usually can be decomposed into sub-goals. For example, if an investigation includes three experiments, then the goal of the investigation can be decomposed into a set of sub-goals, and each of them constitutes a goal of the experiment. A goal often appears in the Abstract or Introduction sections of the paper but it is not uncommon for sub-goals to be mentioned in the Results and Discussion sections. The GSC annotation scheme makes no distinction between goals and sub-goals. A goal is usually introduced through an infinitive or variations on characteristic phrases such as *'we aim to...', 'we investigate, examine',* etc.

Example 1:

> *"Our aim was to test whether this reaction could be enhanced in presence of light."*

This sentence clearly states an aim, "to test a particular reaction" and should therefore be labelled as <Goal> and given a new concept ID.

Example 2:

> **"***Computer modelling techniques have been employed to qualitatively and quantitatively investigate the dehydration of hydroxyapatite to oxyapatite and the defect chemistry of calcium-deficient hydroxyapatite, where a number of vacancy formation reactions are considered."*

This sentence mentions the types of techniques used to achieve a goal in this investigation, but it is primarily about the goal: "*to… investigate the dehydration of hydroxyapatite to oxyapatite*". Therefore, it should be annotated as <Goal> and given a new concept ID.

Example 3:

> *"Therefore, an important aspect to this study is the implementation of theoretical methods capable of calculating ground and excited state energy level differences, which together with knowledge of symmetry-allowed transitions can be used to reproduce (stick) spectra."*

This sentence talks about a goal of the study and should be annotated as <Goal> with a new concept ID.

Example 4:

> "We first discuss the formation of oxyapatite from hydroxyapatite, followed by various calcium-deficient hydroxyapatite structures, where the calcium vacancies are charge-compensated by a range of further defects, namely hydroxy group vacancies, phosphate vacancies and substitutional carbonate groups."

This sentence presents some of the goals of the study in order and should be annotated as <Goal> with a new concept ID.

Example 5:

"First demonstration of implementing the method for InP and GaAs is reported ."

Again this sentence mentions a goal of the investigation, namely reporting the outcome of implementing the method for InP and GaAs.

Example 6:

"*To demonstrate the generality of the approach and its applicability to additional semiconductors, we made first attempts to synthesize InP and GaAs nanorods*."

Again this sentence mentions a goal of the investigation, namely synthesising InP and GaAs to show the generality of their approach. In actual, fact this sentence rephrases the goal of the sentence in example 5, so the two of them receive the same concept ID.

Example 7:

"*We will show that there is a substantial effect of light on the NO2 uptake kinetics on organic surfaces and on the formation of nitrous acid*".

This sentence mentions a goal of the investigation, which is namely to confirm the initial hypothesis about the effect of light on NO2 uptake kinetics.

Example 8:

"*In this study we have used computer simulation techniques to investigate a number of defects in the hydroxyapatite structure*".

This sentence mentions the goal of the investigation, which is namely to invesstigate defects in the hydroxyapatite structure using computer simulation techniques. This <Goal> sentence should have the same concept ID as the sentence in example 2.

5.5 <Hypothesis>

A hypothesis is a statement that has not been yet confirmed rather than a factual statement. A hypthesis can be expressed as a possible explanation, decision, prediction and proving or disproving a hypothesis is one of the objectives of an investigation
. A hypothesis can refer to an investigation, experiment or model. A hypothesis is rarely stated explicitly in text but can usually be inferred from the goals and objects of an investigation. If there is mention of what authors aim to prove in their investigation or experiment, then mark the sentence containing it with the GSC <Hypothesis>.

Example 1:

> *'Photoenhanced uptake of gaseous NO2 on solid organic compounds: a photochemical source of HONO'?*

From this title/question one can infer that the hypothesis of the investigation is the following: *'Photoenhanced uptake of gaseous NO2 on solid organic compounds is a photochemical source of HONO'* and that the main goal of the investigation is to test this hypothesis. This title/sentence should be assigned the GSC <Hypothesis>.

Example 2:

> *'Therefore, we propose that NO2 is reduced in the present system by reductants produced during the photosensitized oxidation of phenols.'*

Here there is an intermediate hypothesis of the investigation which the authors proceed to try and prove. This sentence should also be annotated as a <Hypothesis>.

The same holds for examples 3-9 below. They are all intermediate hypotheses of the investigation, phrased as statements with a level of uncertainty that the authors proceed to test. Notice the use of the conditional and verbs such as "suggest", "expect", "propose", "may be", "could be". A level of uncertainty is sometimes expressed in a <Conclusion> as well, but the main difference is that there is no further testing of a statement which qualifies a <Conclusion> within the scope of the current investigation. Unlike a <Conclusion>, which is a final outcome of the investigation, a <Hypothesis> is part of the objectives of the investigation, what it tries to prove or gain knowledge about.

Example 3:

"We suggest that these particles are mainly composed of indium."

Example 4:

"Concerning the mechanism of formation of the big InAs particles, we suggest this arises from large indium droplets formed upon the initial injection. The droplets catalyze rapid growth of the big InAs particles while the indium from the droplet reacts with arsenic to form the semiconductor."

Example 5:

"Bulk silver has a lower melting point (962 °C) as compared to Au, therefore larger silver nanoparticles are proposed to be applicable as catalysts for the synthesis."

Example 6:

"We suggest that the shell in this structure grows by a catalytic mechanism instead of an epitaxy mechanism active in growth of semiconductor core/shell nanocrystals[6,41].The catalytic mechanism may be activated in this case by dissolution of the reaction precursors within a surface layer of the Ag nanoparticles."

Example 7:

"InAs provides a very interesting and different situation as the Bohr radius is much larger –35 nm, and significant length dependent effects are to be expected."

Example 8:

"An additional mechanism for the decreasing intensity of the PL in longer rods, could be the presence of an increased number of potential surface trap sites that provide non-radiative decay routes for the excited carriers."

Example 9:

"If traps were indeed the determining cause for the length dependence of the PL intensity, one may expect at low temperature to observe strong red-shifted PL at shorter wavelength and the trapped emission would dominate the signal[47]."

5.6 <Method>-<New>

A method of an investigation is the means by which the authors seek to achieve a goal of the investigation. A research method is a way to solve a scientific task "based upon or regulated by science, as opposed to mere traditional rules or empirical dexterity" [OED]. An experimental method is a special case of <Method>. All sentences which directly describe experiments conducted within the context of an investigation should be marked as <Experiment>. Sentences which describe methods, other than experimental methods (i.e. analytical, theoretical) should be annotated as <Method>. Sentences which provide general, not direct or specific, information about experiments of an investigation should also be marked as <Method>. Sentences mentioning experiments which are not part of the current investigation should be marked as <Method>.

In the case where an investigation is about a method, for example if it proposes a new method and studies characteristics of the latter then the method in question is not "the means by which the authors seek to achieve a goal of the investigation", but an object of the investigation. Under such circumstances a sentence about such a method should be assigned the label <Object>.

A method can appear in any of the sections of the paper but is usually more prominent in the Abstract, Methods and Result and Discussion sections. However, when the disadvantages and advantages of pre-existing methods are stated, they tend to appear in the Introduction as well. This is less the case with the mention of new methods, developed for the purposes of the current investigation.
A method can have advantages and/or disadvantages. If these are encountered, they should be annotated as such and linked to the method to which they refer by the same concept ID. More details about this are provided in the description of the annotation interface. If the description of a method is confined to one sentence, and the latter talks about both the method and its advantages/disadvantages, then the label to assign is <Method> with property values <New> or <Old>, depending on the case. It is more important to indicate whether the method is <New> or <Old>, than record its advantages/ disadvantages.

Example 1:

> *"A method was developed for the propargylamine modification of negatively charged dye labels and involves the addition of 6 modified nucleobases at the 5'-terminus close to the dye label".*

This sentence talks about a new experimental technique but it only provides a generic description of this technique (or method). It should be annotated as <Method>.

Example 2:

> *"To overcome this problem we use the tetramine, spermine as an aggregating agent. Firstly, it interacts very strongly with the phosphate backbone of DNA and neutralises the charge, thus reducing the repulsion between the negatively charged DNA and the negatively charged nanoparticles. Secondly, spermine forms aggregates of silver nanoparticles to provide the roughened metal surface which is required to give surface enhancement.*

The first sentence in Example 2 should be marked as <Method> - <New> with a new concept ID. The second and third sentences talk about advantages of the method and both of them should be annotated as <Method> - <Advantage> with the same concept ID as the first sentence.

More examples of <New>-<Method> can be found below:

Example 3:

"*The perfect and some of the defective apatite lattices were modelled using both density functional theory (DFT) methods and interatomic potential-based simulation techniques."*

Example 4:

"*The basic approach used here for rod synthesis combines the principles of colloidal growth of high quality semiconductor nanocrystals, with the one-dimensional growth achieved via the VLS or more precisely, the SLS mechanism."*

Example 5:

"*Indium nanoparticles were prepared in-situ by reduction of InCl3 with NaBH4*."

Example 6:

"*The growth takes place via the solution–liquid–solid (SLS) mechanism where proper precursors are injected into a coordinating solvent."*

Example 7:

"Centrifugation is used for separation of rod-fractions with different lengths."

Example 8:

"The rods are structurally characterized using transmission electron microscopy (TEM), powder X-ray diffraction (XRD) and energy dispersive X-ray spectroscopy (EDS)".

5.7 <Method>-<Old>

An existing or old method is defined in the same way as a new method in Section 5.6, as the means by which the authors seek to achieve a goal of the investigation. A research method is a way to solve a scientific task "based upon or regulated by science, as opposed to mere traditional rules or empirical dexterity" [OED]. The main difference between an old and a new method is that an existing method is more likely to be found in the literature review section and to be accompanied by citations. If it is not obvious that the method used is a new method it is safer to assume that it is an existing method.

Example 1:

"Previously we synthesized a diene tagged oligonucleotide then conugated a benzotriazole dye maleimide Diels Alder cycloaddition".

In this sentence the authors refer to a method they followed in previous work of theirs.

Example 2:

"In recent years a number of ingenious methods for detecting specific DNA sequences have been reported. The most widely used are based on fluorescence detection during PCR amplification and are closed tube homogenous assays. These include Molecular Beacons,[1] Taqman[2] Scorpions[3] and Hybridisation Probes[4]"

The first stentence "*In recent (..) reported*" should be marked as <Background> since it acts as a precursor to a sentence introducing "pre-existing"/ "old" methods. The second sentence contains the description of GSC <Method> indicated by words "*the most widely used*". This sentence should be marked as <Method> with property value <Old> as the method mentioned is "the most widely used" and not attributable to the authors' current work. It should be given a new concept ID. The 3rd sentence should be marked as <Method> with property old and the same concept ID as the 2nd sentence.

Example 3:

"To date, two complementary experimental techniques have emerged. Using electrospray, a number of groups have been successful in generating a range of doubly and triply charged complexes[5-9]. This approach appears to work particularly well for structural studies of biologically-related ligands;[7] however, Posey et al[8] and Thompson et al[9] have also demonstrated that the technique is capable of yielding information on the electronic spectroscopy of doubly charged complexes."

The first sentence *"to date (..) emerged"* is only an indicator that the following sentences describe "experimental techniques", or methods, in our terminology. It should be annotated as <Background>. The second sentence *"Using (..) complexes[5-9]"* should be labelled as <Method> with the property value <Old> and a new concept ID. An indication, that it is already a known method is the references. The 3rd sentence "*This approach (..) charged complexes*" describes an advantage of the previous method, and should be annotated as <Method> with the property <Advantage>. In order to record that it is the same method as in the previous sentence, one should assign the same concept ID. If the description of a method is confined to one sentence, and the latter talks about both the method and its advantages/disadvantages, then the label to assign is <Method> with property values <New> or <Old>, depending on the case. It is more important to indicate whether the method is <New> or <Old>, than record its advantages/ disadvantages.

5.8 <Model>

A sentence which describes a theoretical model or framework used or formed within an investigation as well as how the model is represented should be annotated with the label <Model>. It can be expressed as a system of equations, a set of principles, the design of an object, or a set of assumptions. A model will be normally described in the Methods or Results and Discussions section. In experimental papers there are few (if any) instances of <Model> unlike in the case of theoretical papers, where there is an abundance of descriptions pertaining to theoretical settings and assumptions.

Example 1:

"The latter techniques are based on the Born model of solids,[30] which assumes that the ions in the crystal interact via long-range electrostatic forces and short-range forces, including both the repulsion and van der Waals attraction between neighbouring electron charge clouds, which are described by simple parameterised analytical functions. The electronic polarisability is included via the shell model of Dick and Overhauser,[31] where each polarisable ion, in our case the oxygen ion, is represented by a core and a massless shell, connected by a spring."

The first sentence "*The latter (..) functions*" appears just after the mention of the theoretical methods being used in the current investigation. This sentence mentions the theoretical framework for the methods and should therefore be given the label <Model> with a new concept ID. The 2nd sentence "The electronic (..) spring." adds to the description of the model and should also be annotated as <Model>, with the same concept ID.

Example 2:

"A triple-ζ Slater-type-orbital (STO) basis set with a polarisation function (TZP) was used to describe the valence electrons of each atom: Cu (3s 3 p 3 d 4 s), Ag (4 s 4 p 4 d 5 s), C (2 s 2 p) and N(2 s 2 p), and electrons in lower shells were treated with the frozen core approximation."

This sentence describes how the valence electrons of each atom where represented. A system of representations is considered as a model. This sentence also can be viewed as the description of a theoretical framework or a method. In any case, GSC <Model> has a higher priority than GSC <Method> (see Section 6). Therefore this sentence should be annotated as <Model>.

Example 3:

"The process can be described by the following two reactions: 2InCl3 + 6NaBH4 → 2In + 6NaCl + 3B2H6 + 3H2, InCl3 + As(Si(CH3)3)3 → InAs + 3(CH3)3SiCl."

This sentence describes a theoretical framework in terms of two reactions and should therefore be annotated as <Model>.

Example 4:

"*The bulk exciton Bohr radius, a0, is a useful ruler for distinguishing between different quantum confinement regimes. In the case where a critical dimension d of a semiconductor nanostructure is smaller than a0 strong quantum confinement occurs, while weak confinement is defined for cases where a0 ≪ d.*"

This sentence describes a theoretical model, the Bohr radius, and how this can be employed to draw conclusions about quantum confinement. It should be annotated as <Model>.

Example 5:

"*The latter techniques are based on the Born model of solids,[30] which assumes that the ions in the crystal interact via long-range electrostatic forces and short-range forces, including both the repulsion and van der Waals attraction between neighbouring electron charge clouds, which are described by simple parameterised analytical functions. The electronic polarisability is included via the shell model of Dick and Overhauser,[31] where each polarisable ion, in our case the oxygen ion, is represented by a core and a massless shell, connected by a spring. The polarisability of the model ion is then determined by the spring constant and the charges of the core and shell.*"

The first sentence describes a theoretical model, the Born model of solids, while the following two sentences give extra information abou the same model. All three sentences should be annotated as <Model> with the same concept ID.

Example 6:

"The lowest-energy structure of this defect in a (1 × 1 × 1) simulation cell had the water molecule located in the channel with both hydrogen atoms pointing towards the oxygen atom and the inter-hydrogen mirror plane along the hexagonal c-axis, while the other structure we considered had one of the hydrogen atoms pointing towards the oxygen atom with the OH bond along the hexagonal c-axis, but with the other hydrogen pointing away from the c-axis at an angle of 75°. A schematic representation of both starting configurations is shown in Fig. 3."

The first sentence describes the theoretical model/structural representation used in a simulation while the second sentence points to its schematic representation. Both sentences should be annotated as <Model> with the same concept ID.

Example 7:

"We have used a simulation supercell, incorporating two hydroxy channels and containing a total of eight hydroxy groups, leading to 25, 50 and 75% substitutions of the hydroxy groups."

This sentence describes the theoretical model/structural representation used in a simulation and should be annotated as <Model>.

Example 8:

"Once the lowest energy positions for the oxygen atoms were determined, the lattice was relaxed to give the optimised geometry of the oxyapatite material."

This sentence describes the theoretical framework used in a simulation and should be annotated as <Model>.

5.9 <Motivation>

A sentence that indicates why the investigation is important should be annotated as <Motivation>. Motivation for an investigation is the stimulus for achieving the goal of the investigation, the reason to carry out the investigation. The motivation is not always explicitly stated in the paper but when present it is usually mentioned in the Abstract and/or the Introduction sections. It is normally phrased as a problem whose resolution is of vital importance for the current investigation.

Example 1:

"As both products of reaction (1) are key players in the processes leading to photochemical pollution, it is essential to understand and identify all sources of tropospheric nitrous acid. Despite many studies, to date the formation mechanism of HONO in the atmosphere is still not completely understood".

The first sentence states the importance of identifying sources of tropospheric nitrous acid and the second sentence emphasizes the lack of such knowledge to date. They should therefore both be annotated as <Motivation> with the same concept ID.

Example 2:

> *'However, standard coupling chemistries using activating agents do not work well with benzotriazole dyes and we have investigated alternative strategies.'*

This sentence presents a problem and the solution as an object of the current investigation. It is therefore clear that the problem presented is a driving force for the investigation. Even though the sentence contains both a motivation and an object, it should be annotated as <Motivation> since the latter ranks higher in the priority of concepts as defined in Section 6.

Example 3:

> *"More recently, apatites have gained additional prominence due to their biological rôle as one of the main constituents of mammalian bones and tooth enamel[8]."*

Again this sentence justifies the current investigation, which looks into hydroxy – and oxy- apatites.

Example 4:

> *"One of the most significant goals in any series of experiments on gas phase transition metal complexes has to be that of studying their ligand field spectroscopy. Such studies would enable the development of a ligand field to be interrogated in the absence of any perturbations from either counter ions or solvents."*

The first sentence "One (..) spectroscopy" indicates that "*studying their ligand field spectroscopy* " is recognised as an important task, and therefore justifies the current investigation. This sentence should be labeled as <Motivation> and should be given a new concept ID. The second sentence elaborates on the possible goals of studying ligand field spectroscopy. Therefore, this sentence should also be annotated as <Motivation> and given the same concept ID as the sentence preceding it.

More examples of <Motivation> can be found below:

Example 5:

"The mechanism of this source remains unclear."\

This sentence presents a problem that constitutes a motivation for the current investigation,

Example 6:

"However, as biological apatite is often both calcium-deficient and rich in carbonate, we clearly need to include this combination of defects in our calculations to consider its possible occurrence in the apatite mineral."

This explains a problem arising that influences the tasks to be faced by the current investigation and it should therefore be annotated as <Motivation>. One could perhaps confuse it with <Goal> but the difference between the two is that the first is stated in an indirect way and may or may not include a task followed by its justification whereas the latter clearly states one or more tasks to be addressed by the current investigation.

Example 7:

"In order to calculate the defect formation energy, we also need to know the energy of dissociation of the water molecule, but this reaction requires the second electron affinity of oxygen, which is material-dependent."

Example 8:

"However, defects in the lattice can have a large effect on the elastic constants of a material and it is important for biomaterial applications of apatite minerals that the elastic constants for defective structures are known[79]."

Example 9:

"One of the most important challenges in current nanocrystal research is that of achieving shape control as the properties of these quantum-confined structures can also be modified by the change of shape[13–21]."

Example 10:

"However, the exact mechanism of this photolytic HONO source still remains to be unanswered, e.g. the photolysis frequency of adsorbed HNO3 was reported to be two orders of magnitude faster compared to the gas and the liquid phase, which is still unclear[38,39]."

5.10 <Object>

An <Object> of investigation is an entity or series of entities about which the authors seek to gain new knowledge by means of the investigation. It is the principal entity of an investigation. There can be several objects of investigation in a paper. The word 'object' should not be taken literally as pertaining to some physical object necessarily. An object of an investigation is regarded as a role played by some entity (a role holder). An object of investigation can be a physical object, but also a process, a method, or an abstract entity. The word "object" then reflects the relation between this entity and the investigation in the sense that the former denotes what the investigation is about.
An investigation could then be about developing a new method or testing a new method, and this method is playing the role of an object of investigation. Therefore sentences that describe methods which constitute the topic of an investigation should be labelled as <Object>.

An object of investigation often appears in the Abstract or the Introduction section of a paper as well as the Conclusion section. However, it is not uncommon for objects of investigation to be mentioned in the Results and Discussion sections as well. An object of investigation is often introduced as the object of verbs such as '*report', 'discuss', 'investigate', 'study'* or as subjects of these verbs in passive voice. An object of investigation can also be signalled by an *'is a'* expression, providing a definition of the object.

Similar to the <Method> GSC, an <Object> of investigation can have advantages and/or disadvantages. If these are encountered, they should be annotated as such and linked to the object to which they refer to by using the same concept ID. Note that we are only interested in new objects of investigation. Therefore, an <Object> of investigation when selected should have the property <New> by default.

An <Object> can be distinguished from a <Goal> as the latter represents the task or tasks to be accomplished by the investigation whereas the former consists of the particular entities being studied by the investigation. There are normally many more instances of <Object> in a paper than instances of <Goal>.

Example 1:

> *'The effect of light on the uptake kinetics of NO2 on various surfaces taken as proxies for organic surfaces encountered in the troposphere was investigated.'*

The phrase **"***was investigated"* clearly points to an object of the current investigation. The sentence should be annotated as <Object>.

Example 2:

> *'The uptake of NO2 in the dark was also studied on various pure organic surfaces'.*

The phrase **"***was also studied"* points to another object of the current investigation. The sentence should be annotated as <Object>.

Example 3:

> *"Some of the complexes have been the subject of a study of charge transfer transitions at UV wavelengths,[17] and a preliminary report on d–d spectra recorded for* $[M(py)4]^{2+}$ *complexes has also been presented[18]."*

This sentence mentions a "*subject of a study"*, and one would consider annotating it as <Object>. However, the sentence talks about an object of some other study, previous work, not the one reported in the current paper. Therefore, the particular sentence does not constitute an <Object> of investigation, which as we have seen should always have the property <New>. The sentence in question should be annotated as <Background>.

Example 4:

> *"The uptake of NO2 in the dark was also studied on various pure organic surfaces i.e., on catechol, anthracene, benzophenone, anthrarobin."*

Example 5:

"*Three different types of light-absorbing organic species have been investigated. These are 4-benzoylbenzoic acid (4-BBA) and its sodium salt, 3,7-dihydroxy-2-naphthoic acid (3,7-DHNA) and perylene (PER)."*

Example 6:

"*A third UV-A absorbing species 3,7-DHNA was chosen, as hydroxylated aromatic compounds are known to react with NO2 in aqueous solution or on wetted surfaces to form nitrite or HONO and aryloxy radicals in a dark reaction[23,24,27,34]."*

Example 7:

"*A new approach for the detection of DNA using surface enhance resonance Raman scattering (SERRS) is reported*."
Here a method, the new approach for the detection of DNA, plays the role of object of investigation and should therefore be annotated as <Object>.

Example 8:

"*In this study we report the synthesis of a new class of biomolecular probe called a SERRS Beacon, which uses the detection technique of surface enhanced resonance Raman scattering, SERRS. It offers the option to use fluorescence if desired. SERRS has the advantage that it is a molecularly specific technique allowing sharp fingerprint spectra to be obtained, thus allowing discrimination between components in a sample without the need for separation. In addition, it is an extremely sensitive technique with single molecule detection reported[6,7]".*

Again, "*the synthesis of a new class of biomolecular probe"*, plays the role of object of investigation and therefore the first sentence should be annotated as <Object>. The following three sentences provide advantages of the <Object>, and should therefore be annotated as <Object><Advantage> with the same concept ID as the first sentence.

Example 9:

"A novel approach for synthesis of soluble semiconductor quantum rods using metal nanoparticles to direct and catalyze one-dimensional growth is developed. The method is useful in particular for III–V semiconductors with cubic lattice, where the utilization of surfactant-controlled rod-growth is not easily realized."

Again, "*A novel approach for synthesis of soluble semiconductor quantum rods using metal nanoparticles to direct and catalyze one-dimensional growth"*, plays the role of object of investigation and therefore the first sentence should be annotated as <Object>. The following sentence provides an advantage of the method in the previous sentence, which plays the role of object, and should therefore be annotated as <Object><Advantage> with the same concept ID as the first sentence.

Example 10:

"Here we extend this approach and investigate the use of different metals as the catalyst including different forms of Au clusters, Ag and In nanoparticles."

Example 11:

"We also study the effect of reaction temperature, duration and concentration of precursors on the reaction products."

Example 12:

"We report here a systematic study of the effect of growth conditions in this reaction scheme."

Example 13:

"Au101 clusters used in the synthesis here are smaller than the nanoparticles (diameters 1.5 nm and 2 nm, respectively), and have a narrower size distribution."

Here the object is the Au101 clusters and the sentence should be annotated as <Object>.

Example 14:

"The preparation of length-controlled InAs rods allows us to conduct a study of the size-dependent quantum confinement effect, and in particular the transition from a 0D dot to a 1D quantum wire."

Here the sentence includes the object of investigation "*a study of the size-dependent quantum confinement effect"* and should therefore be annotated as <Object>.

Example 15:

"The present study reports a detailed, atomic level computational study of the incorporation of various defects which may contribute to the calcium deficiency in the apatitic phase of bone, including OH vacancies, PO4 vacancies and substitutional carbonate defects. Computational methods are well placed to calculate at the atomic level the defect formation energies and structures of the calcium-deficient apatite materials, which can be difficult to obtain experimentally, especially for natural bone material[12,13]."

Again, the computational study in the first sentence, plays the role of object of investigation and therefore the sentence should be annotated as <Object>. The second sentence provides an advantage of the method in the previous sentence, which plays the role of object, and should therefore be annotated as <Object><Advantage> with the same concept ID as the first sentence.

Example 16:

"*We first investigated the replacement of OH groups in the hydroxyapatite structure by an oxygen ion and water molecule, in effect the products of the dissocation of one hydroxy ion and the addition of the resulting proton to the second hydroxy ion."*

Example 17:

"The experiments employing iodide as the electron donor were conducted to investigate an alternative route to produce the ketyl radical of 4-BBA62,[63]. "

Here the sentence contains both a mention of method (generic desction of experiments) and an object ("*alternative route to produce the ketyl radical of 4-BBA62").* Since the priority of concepts rates

<Object> higher than <Method> this sentence should be given the annotation <Object>.

Example 18:

"The SERRS Beacon is a dual labelled probe with a fluorophore at one end and a different dye at the other. In conventional Beacons the second dye is a quencher such as DABCYL but in this case we have used a specifically designed benzotriazole dye that is capable of immobilising the oligonucleotide probe onto a silver metal surface."

The first sentence defines the SERRS Beacon (note the "is a" expression), which is an object of the current investigation, and should therefore be annotated as <Object>. The second sentence talks about properties of the SERRS Beacon that give it advantages over conventional beacons and should be labelled as <Object><Advantage> with the same concept ID as the first sentence.

Example 19:

"The design and synthesis of a new class of DNA probe has been reported as has a new method adding a surface seeking dye label to DNA."

This sentence includes two objects of investigation; "The design and synthesis of a new class of DNA probe" and a "a new method adding a surface seeking dye label to DNA", where the latter plays the role of object of the investigation. Therefore the sentence should be annotated as <Object>.

5.11 <Observation>

An <Observation> encompasses the data/phenomena recorded within an investigation, as well as how the data are represented (e.g. in the form of links to data bases, as tables, figures, diagrams). An experimental observation is a direct observation of nature or a set of values resulting from measurements, prior to interpretation. For example, measurements or descriptions of the direct outcomes of an experiment or a method qualify as observations. Observations are often signalled by the verbs *'show', 'observe', 'see', 'detect'* and the inclusion of quantities, data, figures and measurements as well as their numeric or non-numeric description.

Example 1:

> *"Ligand field spectra have been recorded in the gas phase for the two series of complexes containing either Cu(ii) or Ag(ii) in association with pyridine."*

This sentence talks about an outcome of the investigation. It should be annotated as <Observation>, because it gives a record of evidence ("*spectra have been recorded'*), without the involvement of any post-processing by researchers.

Example 2:

> *"Using this detection strategy it has been possible to obtain SERRS signals from 12 commercially available fluorescent labels[8,23] examples of the spectra obtained are shown in Fig. 1. The spectra shown were all obtained using 514.5 nm laser excitation, however, in the case of Cy5, Cy5.5, Cy3.5 and BODIPY TR-X lower detection limits were actually obtained when 632.9 nm laser excitation was used since the absorbance maxima of these labels are closer to this excitation wavelength than 514.5 nm."*

The first sentence, "*Using (..) Fig. 1*.", is an example case where the sentence should be annotated as <Observation> even though the actual observations "*SERRS signals"* are included in the figure it refers to. The second sentence, "*The spectra (..) 514.5 nm*." provides experimental details and explanations about the spectra mentioned in the first sentence and should be labelled <Observation> with the same ID.

Example 3:

> *"Only a physical adsorption was observed i.e., the surface exposed by the glass flow tube was rapidly covered and saturated by NO_2 molecules".*

The phrase "*was observed*" indicates the description of observations. The sentence should be annotated as <Observation>.

Example 4:

> *"When the organic surface was exposed to gaseous NO_2, we first observed a rapid decrease of the MS signal at m/z = 46 (i.e., corresponding to the NO_2 molecular mass) which recovered to its original level within a few minutes".*

The phrase "*was first observed*" again indicates the description of the observations. The sentence should be annotated as <Observation>.

More examples of <Observation> can be found below:

Example 5:

"An example of the SERRS spectra obtained from the benzotriazole azo dye labelled oligonucleotide is shown in Fig. 2."

Example 5:

"The fluorescence quenching effect improved from 88% for DABCYL/FAM to 96% for BT dye/FAM and then to 98% when the silver nanoparticles were added."

This sentence includes the direct effect of using different dyes in a molecular beacon and should be annotated as <Observation>. Even though both experiments and observations often incorporate descriptions of experimental conditions and configurations, one can easily distinguish between them since <Observation> refers to the outcome of the experiments whereas <Experiment> is reserved for the description of the experimental procedure and does not usually include outcomes.

The same applies to examples 6 and 7 below.

Example 6:

"The DABCYL/FAM pair did not show any change when silver was added."

Example 7:

"*Under these conditions (1.2 × 10–7 M) no fluorescence could be detected using a Cary Eclipse fluorimeter, a Stratagene MX4000 fluorescence plate reader and a Renishaw Raman spectrometer due to the highly efficient quenching by the silver nanoparticle."*

Example 8:

"*Fluorescence is now emitted and the SERRS signals change to show the distancing of the fluorophore from the silver surface."*

Example 9:

"It is seen that indium particles with diameters ranging from a few nanometers to ~20 nm were obtained."

Example 10:

"These products can be identified in the powder X-ray diffraction pattern in Fig. 4b. Narrow diffraction peaks are observed as the signal is dominated by the large particles."

These two sentences should be annotated as <Observation> with the same concept ID.

Example 11:

"*Then, the solution was centrifuged at 4000 rpm (rcf ~1800g) for 10 min, and the supernatant contained a fraction of nanorods with an average length of ~25 nm and diameter of 3–4 nm as shown in Fig. 3c."*

Example 12:

"At lower growth temperatures more kinks are observed in the separated rods. The yield of rods in this case is low."

These two sentences should be annotated as <Observation> with the same concept ID

Example 13:

"Nanorods were only obtained when the ratio was as high as 9 although a large fraction of the rods are kinked and their yield was low."

Example 14:

"TOPO provides higher reaction temperatures, resulting in immediate formation of rods upon injection of the precursors."

5.12 <Result>

Results are statements about the outputs of an investigation in the form of facts, or factual statements, supported by the observations and other results. Results of the investigation are obtained through the interpretation or comparison of observations. Observation provide the evidence for <Result>, which constitutes one step further in terms of inference and abstraction. This is how one can distinguish between <Observation> and <Result>, since the former provides the basis of the latter.

Example 1:

> *"Where comparisons are possible, the gas phase spectra match those recorded in the condensed phase; however, for Ag(ii) systems the results differ in interpretation."*

This sentence should be annotated as <Result>, because this outcome includes some processing of directly observed data ("*comparisons", "results differ in interpretation*").

Example 2:

> *"The results from these experiments (Table 2) show that the phototransformation of NO2 -> HONO is also effective when the electron donor is not a phenolic compound."*

The sentence mentions data directly observed (Table 2), but also their interpretation, that is, the process or model reflected by the data. Therefore, the sentence should be annotated as <Result>. Whenever there is a case of competing GSCs within a sentence, one should always consult the GSCs priority list in Section 6.

Example 3:

> *At all values of n except 2, absorption was observed at wavelengths between 450 and 650 nm, which resulted in the following photofragmentation channels:*
>
> $[Ag(py)_3]^{2+} \rightarrow Ag(py)^{+} + py^{+} + py$
>
> $[Ag(py)_4]^{2+} \rightarrow Ag(py)_2^{+} + py^{+} + py$
>
> $[Ag(py)_5]^{2+} \rightarrow [Ag(py)_4]^{2+} + py$
>
> $[Ag(py)_6]^{2+} \rightarrow [Ag(py)_4]^{2+} + 2\ py$

This is another example case where a sentence mentions both observations (the wavelength range) and resuls (the equations which represent the photofragmentation channels to be inferred from the the observations). Because <Result> is higher on the GSC priority list than <Observation>, this sentence should be annotated as <Result>. Also note that even though this sentence includes a set of equations, the latter constitute an outcome and not the theoretical framework for the current approach, therefore the sentence should NOT be annotated as <Model>.

Example 4:

> *"Although an accurate measurement of absorption cross sections was not possible, primarily due to the uncertainty of the length of the overlap between the laser and the cluster beam in the second field free region (2nd ffr) of the mass spectrometer, an estimate of the extinction coefficient for* $[Cu(py)_4]^{2+}$ *was obtained."*

"An estimate of" implies some calculation or inference drawn from the observations. Thus, this sentence should be annotated as <Result>.

More examples of <Result> can be found below:

Example 5:

"However, with anthracene, NO2 never reached again its initial concentration, within our experimental conditions, but reached a plateau which is an evidence that at long enough gas/solid contact times, the uptake is driven by a slow dark reaction. However, this was observed to slow down at long reaction times i.e., the plateau is not totally time independent, which means that there is a deactivation of the surface during the course of the experiment."

These two sentences both interpret evidence and experimental observations from the same reaction and they should be therefore annotated as <Result> with the same concept ID.

Example 6:

"For the synthesis of InAs nanorods with indium catalyst, we found that both TOP and TOPO can be used as solvents, but TOPO has a higher boiling point than TOP (360 vs. 300 °C), and thus enables a higher reaction temperature resulting in better crystallinity and yield of the products."

This sentence provides a comparison between the performance of two solvents and a more abstract interpretation of what this comparison means. It should be thus annotated as <Result>.

Example 7:

"We could estimate that about 80% of the product is in the form of small InAs nanodots."

This is an estimation about observed data and it should therefore be labelled as <Result>.

Example 8:

"With such short reaction time the growth of large indium nanodroplets can be mostly avoided, enabling the use of a lower ratio that results in a higher yield of rods. A ratio of ~4.5 was found to be optimal in our typical synthesis conditions."

These two sentences interpret the observations preceding them, stemming from experiments in nanorod synthesis in the current investigation. The second sentence provides an estimate fpr an optimal ratio. The two sentenes should both be annotated as <Result> with the same concept ID.

Example 9:

"The synthesis in this case yields InAs quantum rods with a uniform diameter but broad length distribution, along with byproducts including a small fraction of InAs quantum dots, InAs nanowires, and gold particles."

This sentence is given in lieu of any observations, so it summarises all the observations from certain experiments. This sentence should also be annotated as <Result>.

Example 10:

"These dots can be identified by TEM to have a core-shell structure with a gold core and InAs shell."

Example 11:

"This indicated that the FAM was removed from the silver surface."

Example 12:

"In the closed SERRS Beacon (Fig. 3) the FAM signals dominate since FAM gives stronger signals than the benzotriazole dye by approximately two orders of magnitude."

Example 13:

"The spectrum is clearly different to the fluorescent dye spectra in Fig. 1 due to the different molecular structure and offers additional opportunities for increased multiplexing."

Example 14:

"This means that the organic surfaces were very rapidly saturated with NO2 molecules, as on the bare glass surface."

6 Priority of Concepts

GCSs are defined in such a way there is no intersection in their semantics (meaning). However, a natural language is not as succinct and precise, which makes it sometimes difficult to identify the best GSC with which to label a sentence.

Even though these guidelines have been written to make the distinction between different GSCs as clear as possible, there will be cases where it will be hard to choose a GSC to assign to a sentence. Indeed, there will be sentences including more than one GSC concept. However, the current annotation scheme only allows one GSC annotation per sentence. In cases where a sentence potentially includes more than one GSC one would need to know which concept to give priority to. Consider for example the sentence:

Example:

> *"In this study we report the synthesis of a new class of biomolecular probe called a SERRS Beacon, which uses the detection technique of surface enhanced resonance Raman scattering, SERRS."*

This sentence contains both an <Object> (the synthesis of the new biomolecular probe) and a <Method> (SERRS). We would give priority to the <Object> label in this case, as the method is more likely to be repeated later in greater detail, in the Methods section whereas goals and objects of investigation have fewer mentions in the text.

Please have the following ranking as a guide of preference for concept assignment, from **higher** to **lower** priority:

<Hypothesis>
<Goal>
<Motivation>
<Object>
<Conclusion>
<Result>
<Model>
<Experiment>
<Method>
<Observation>
<Background>

Note that the order of preference is based on a combination of GSC importance and expected rarity of their occurrence in the paper. For example, we do not expect many instances of hypothesis in the text, so if there is a mention of it in the paper, we would like to give it a higher priority than another potential concept which may be included in the same sentence.

If a sentence contains both a <Method> and an <Experiment>, then priority should be given to the latter. However, if one determines that the distinction between the two is not clear in a particular

sentence, then it is safer to annotate that sentence as <Method> since this is a more general concept than <Experiment>.

7 Conflict Resolution

<Object> vs <Goal>

Goals and objects of investigation are obviously related but they can be distinguished by virtue of the fact that an object is an entity of some sort (e.g. method, process, pattern) playing the role of an 'object' whereas a goal is the task which the investigation seeks to accomplish. In order to better understand how objects and goals are related and differ, let us assume that we have correctly identified objects of an investigation: object 1, object 2, object 3. The sentences describing these objects should all be labelled as <Object>. Goals of an investigation will be statements expressing what the authors intend/ed to do with these objects. Examples of goals are "to find out how object 1 is different from objects 2 and 3", "to investigate what characteristics of object 2 are", "to design object 3". More precisely: an object of an investigation is usually present in the sentences stating goals. However, the main focus of sentences describing goals should be "what to do" with these objects, and not "what" these objects are.

Example 1:
'The design and synthesis of a new class of DNA probe has been reported'.

In this example, *'new class of DNA probe'* is an object of investigation. If this sentence were to be given the <Goal> GSC, we would expect it to be phrased along the lines of: *'In this stydy we aim to design and synthesise a new class of DNA probe'.*

Example 2:

"A new approach for the detection of DNA using surface enhance resonance Raman scattering (SERRS) is reported."

<Object> is the most appropriate category for this sentence. The clue here is the nominalisation, "the new approach". This new approach or method can be viewed as an object with potential advantages and disadvantages. If this had been phrased as : "Using SERRS to detect DNA." it would have been labelled as <Goal>.

<Object> vs <Method>

The confusion between the labels <Object> of an investigation and <Method> arises if an object of an investigation is a method. It is a common situation: there are many investigations that propose new methods and study their properties. The best way to avoid such confusion and miss-labelling is to have a clear structure of the investigation in mind, that is, know what the investigation is about (<Object>), what it aims to achieve (<Goal>), how and by what means (<Method>) and what the final output is (<Observation>, <Result>, <Conclusion>). This is why it is important to read each of the papers and identify the structure of the investigation before proceeding with the annotation.

<Object> vs <Conclusion>

The difference here is in deciding whether the content of the sentence constitutes a target or an outcome.

Example:

"This paper shows how careful design of a biomolecular probe coupled with SERRS has created a new method of analysis capable of single base mismatch detection at ultra low concentrations with potential for increased single tube multiplexing."

<Conclusion> is the preferred choice here, since the sentence summarises the main contribution of the paper. The authors are not talking about what they aim to show or what they are investigating, this is the actual outcome of their investigation.

More examples of <Object> are available in Section 5.10.

<Result> vs <Conclusion>

Results are proven facts which are easy to check, because they constitute direct interpretations of some piece of evidence. They are based on the observations and other proven results reported in a paper. The confidence that observations are true facts is high. In contrast, conclusions are generalisations, inferences, even suggestions based on the outcomes of the investigation. They are not simple facts, and they are more difficult to check as more complicated steps of inference are involved in the production of conclusions. They are not necessarily always correct.
However, they are important as they summarise the investigation and its contribution to knowledge. There are many fewer instances of <Conclusion> in a paper compared to instances of <Result> or <Observation>.

Example 1:
"No significant photoenhancement of the reaction of NO2 with the clean glass surface was observed under humid conditions, in good agreement with previous studies, which showed that reaction (2) is not photoenhanced [39,40]."

This is a comparison between the observations of the current study and the conclusion of previous studies ("reaction (2) is not photoenhanced [39,40]"), which allows the current study to arrive to the same conclusion. Therefore, the sentence contains both the evidence for a conclusion and the conclusion itself and according to concept priority, it should be given the label <Conclusion>.

Example 2:
"Based on the experimental evidence presented in this study, photoinduced conversion of NO2 into HONO has been observed on organic films, which exceeds the rate of the dark reaction by one order of magnitude (see Fig. 5) and that of reaction (2) by even more than one order of magnitude under our experimental conditions."

This sentence should be marked as <Result> as it contains comparisons between observations and infers an outcome from experimental evidence. <Conclusion> is not the appropriate category, as

this sentence conveys a direct result from the observations and not a high level interpretation of the experiment. However, there is a fine line distinguishing the two. Yet the distinction between <Conclusion> and <Result> is important as they do not have the same meaning.

<Result> vs <Observation>

Observations are direct recordings of natural phenomena, or raw data, without any interpretation by the researcher (which e.g. might introduce inconsistencies). Observations can be reused for other investigations. The confidence that observations are true facts is very high. Results are proven facts , which use observation and other proven results as their evidence. They often constitute interpretations or comparisons of experimental evidence.

Example 1:

"Especially organic substrates containing a combination of electron donors, such as phenols, and of compounds yielding excited triplet states, such as aromatic ketones, showed a high reactivity towards NO2."

The sentence should be labelled as "Result" instead of "Observation", because even though the main verb is "show", the content of the sentence does not refer to a direct observation but rather to the interpretation of a series of observations in the experiment.

Example 2:

"However, with anthracene, NO2 never reached again its initial concentration, within our experimental conditions, but reached a plateau which is an evidence that at long enough gas/solid contact times, the uptake is driven by a slow dark reaction."

The sentence contains both the observation ("*However, with anthracene, NO2 never reached again its initial concentration, within our experimental condition"s*) and the result, for which the observation constitutes evidence ("*but reached a plateau which is an evidence that at long enough gas/solid contact times, the uptake is driven by a slow dark reaction*").

In such cases, the sentence should be marked as "Result", according to the instructions in Section 6 for conflict resolution.

Example 3:

"In fact, in this later case the uptake coefficient is no longer time dependent, as the NO2 concentration, when exposed to the irradiated flowtube, decreases rapidly and reaches a plateau after a few seconds."

The same applies to this sentence, which contains both the result ("*In fact, in this later case the uptake coefficient is no longer time dependent*") and the observation that constitutes the evidence ("*as the NO2 concentration, when exposed to the irradiated flowtube, decreases rapidly and reaches a plateau after a few seconds*").

Example 4:

"Here the quantity of NO2 that has reacted is depending "only" on the surface exposed to the gas phase."

The sentence mentions an intermediate result, which is the interpretation of observations on NO2 reactions.

Example 5:

"It was possible to obtain SERRS signals from all 12 of the labels in a quantitative manner allowing detection limits to be obtained which were in the order of ~1 × 10–12 mol dm–3."
"When the detection limits of SERRS, using a commercially available Raman spectrometer, were compared with those obtained using standard quantitative PCR instrumentation, it was found that SERRS was generally at least three orders of magnitude more sensitive than fluorescence[24]."
"The results shown in Table 1 illustrate this."

Most people annotated sentences 1 and 3 as <Result> rather than <Observation>. The distinction between the two is often subtle, but what gives the clue here is that the 1st and the 3rd sentences refer to actual observations whereas the 2nd one is a result, derived from these observations. Te 3rd sentence contains the evidence for the result in sentence 2. The use of the words "results" here is a bit confusing. We want to annotate result as a concept, which may not always coincide with the linguistic use of the term.

Example 6:

"In the closed SERRS Beacon (Fig. 3) the FAM signals dominate since FAM gives stronger signals than the benzotriazole dye by approximately two orders of magnitude."

The sentence a <Result>, as it involves the comparison between signals.

More examples of <Result> can be found in Section 5.12.

<Method> vs <Experiment>

The label <Experiment> should be reserved for sentences which are direct descriptions of experiments conducted in the context of the current investigation. In all other situations, that is, other methods (not experimental), generic description or mention of experiments, experiments that do not form part of the current investigation, the label <Method> should be assigned.

Examples for the distinction between <Experiment> and <Method> can be found in Section 5.3 and Section 5.6

Short Definitions

Sentence: In the current context, a sentence is a meaningful piece of text, containing at least one verb or nominalisation, ending in a punctuation mark (usually '.'). A sentence usually starts with a capital, numeral or italicised character. The annotation interface SAPIENT uses a sentence splitter to detect the boundaries of sentences.

General Scientific Concept (GSC): A GSC is a concept vital to the description of a scientific investigation. They include:

<Conclusion>, <Result>, <Goal>, <Object>, <Method>, <Experiment>, <Observation>, <Hypothesis>, <Motivation>, <Background>, <Model>.

Conclusion: A conclusion of an investigation is a statement summarising the contribution of the investigation, inferred from observations, results, assumptions, and other facts.

Result: Results of the investigation are the set of facts, obtained through the interpretation of observations [EXPO].

Goal: A goal of an investigation is the target state of the investigation where intended discoveries are made, approaches are tested, problems are demonstrated, tasks formulated etc [EXPO].

Object: An object of an investigation is an entity that constitutes a focal point of the investigation.

Method: A research method is a way to solve a scientific task "based upon or regulated by science, as opposed to mere traditional rules or empirical dexterity [OED].

Experiment: A scientific experiment is a procedure which permits the investigation of cause-effect relations between known and unknown (target) variables of the domain. Experimental results cannot be known with certainty in advance [EXPO].

Observation: An experimental observation is a direct observation of nature, the set of values of target variables (or other variables of the domain), "prior to analysis or interpretation". [based on OED].

Hypothesis: A statement about the given information and assumptions as well as the expected conclusions in the context of a study, investigation or model.

Motivation: Motivation for an investigation is the stimulus for achieving the goal of the investigation, the reason to carry out the investigation.

Background: A background fact is a neutral or widely accepted statement about the knowledge domain.

Model: The theoretical framework or a system of representations used or formulated within an investigation.

References

(1) Soldatova, LN & King, RD. (2006) An Ontology of Scientific Experiments. Journal of the Royal Society Interface **3/11**: 795-803.

(2) Soldatova, L., Liakata, M. (2007) An ontology methodology and CISP - the proposed Core Information about Scientific Papers. JISC project report (http://ie-repository.jisc.ac.uk/137/)

(3) Soldatova, L.N., Batchelor, C.R., Liakata, M., Fielding, H.H., Lewis, S. and King, R.D. (2007) ART: An ontology based tool for the translation of papers into Semantic Web format. Proceedings of SIG/ISMB07, pp. 33-36.

(4) OED. The Oxford English Dictionary. (1989) Oxford University Press, 2 Ed.

(5) EXPO. http://expo.sourceforge.net/