Summary Nugget Annotation Guidelines

Table of content

Introduction

The annotations made according to these guidelines support the manual creation process of summaries from multiple sources. In the times of information overload, the task of summarization provides support to information processing. Moreover, these guidelines describe further steps for manual summarization. Please note that the examples are in German, with descriptions translated into English. The next version of the guidelines will contain English examples.

Here, **annotation** describes the process of marking and classifying elements of text. These guidelines describe the annotation of different given texts on one topic. **Nuggets** are text extracts of variable length and annotated in the first step. They contain information that should be used in the summarization of the documents. The following exemplary sentence of a text on the topic of *Open Access* shows a possible annotation of nuggets. However, it shall be stressed that there is no unique annotation of these elements.

> Die Max Planck Digital Library legt eine Analyse zur Transformation des Subskriptionswesens für wissenschaftliche Zeitschriften zu Open Access vor und präsentiert erstmalig quantitative Parameter für diesen Übergang. Sie bestätigen: Die seit Jahren geforderte Öffnung wissenschaftlicher Literatur ohne Mehrkosten ist möglich.
>
> **Example 1 Possible nugget annotation on the topic Open Access (Open Access – Chancen für den Zugang zum Wissen für alle 2007)**

Annotation procedure

The following figure shows the process of annotation from nugget annotation to the finished summaries. The last chapter summarizes the procedure and can be used as a quick look-up device during annotation.
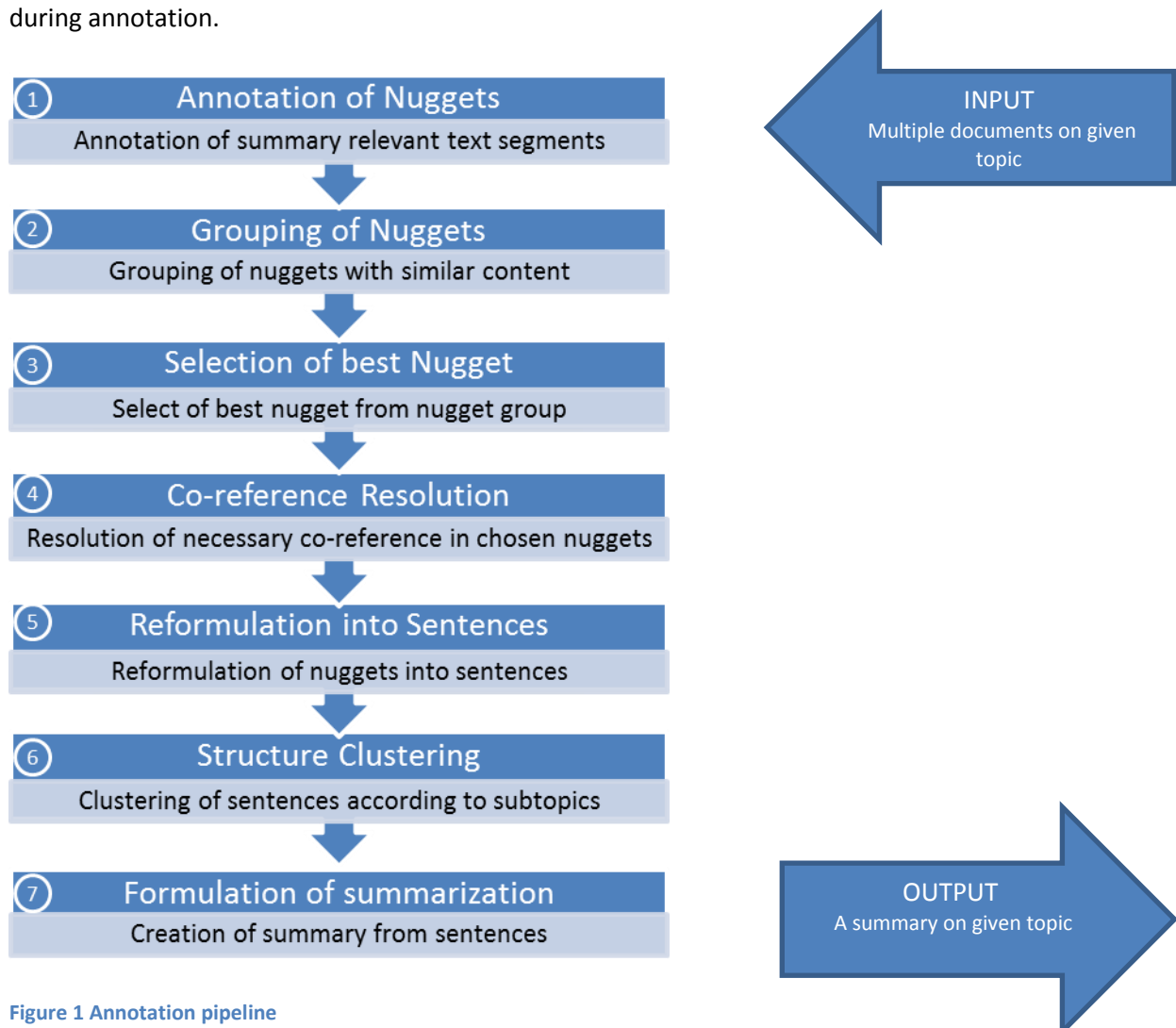
① **Annotation of Nuggets** Annotation of summary relevant text segments	**INPUT** Multiple documents on given topic
② **Grouping of Nuggets** Grouping of nuggets with similar content	
③ **Selection of best Nugget** Select of best nugget from nugget group	
④ **Co-reference Resolution** Resolution of necessary co-reference in chosen nuggets	
⑤ **Reformulation into Sentences** Reformulation of nuggets into sentences	
⑥ **Structure Clustering** Clustering of sentences according to subtopics	
⑦ **Formulation of summarization** Creation of summary from sentences	**OUTPUT** A summary on given topic

Figure 1 Annotation pipeline

First of all, multiple documents on a pre-defined topic are provided. In the first step, the nuggets are marked like it was shown in the introduction. After grouping nuggets with the same content, the best nugget of each group is chosen. Then, all co-references in the nuggets are resolved so that the nuggets become comprehensible without their context. In the next step, the nuggets are reformulated to grammatically correct sentences. Thereafter, the sentences are re-arranged according to a potential structure of a final summary.

> Introduction
>
> Es ist erst ein paar Wochen her, da hatte der ukrainische Präsident Petro Poroschenko seinen großen Auftritt auf internationalem Parkett. Er hielt in der Schweiz eine viel beachtete Rede, im Rahmen des Weltwirtschaftsforums in Davos.
>
> **Example 2 Nugget grouped under Introduction (Bidder 2015)**

In the previous example, general information like "ukrainischer Präsident" (Ukrainian president) being the function of "Petro Poroschenko" (Petro Poroshenko) is clustered as introductory information. In this way, the herein described guidelines not only support the creation of manual summaries, but also provide rules for the structuring of information.

1. Nugget Annotation

What are nuggets?

Information nuggets are also termed *Elementary Content Units*, *Summary Content Units* or Factoids in the context of summaries or question-answer systems (Nenkova et al. 2007).

In these guidelines a nugget is defined as a semantically essential point that is classified as *important* with respect to its relevance in the final summary(Lin, Demner-Fushman 2006).

Nugget criteria

- A nugget should be an appropriate answer to the following question:

  What is important in the context of the given topic?
  Two exemplary nugget annotations on the topic *Open Access* are given in Example 1 and Example 1.1.

  Beschluss
  Die Deutsche UNESCO-Kommission unterstützt das *Open-Access*-Prinzip. Open Access im Sinne dieser Resolution bezeichnet den freien, für die Nutzer im Regelfall kostenlosen Zugang zu mit öffentlichen Mitteln produziertem Wissen.
  Die Deutsche UNESCO-Kommission sieht in *Open Access* eine neue Chance, allen Menschen einen umfassenden und ungehinderten Zugriff auf das mit öffentlichen Mitteln produzierte Wissen zu ermöglichen.
  Example 1.1 Fitting nugget on the topic Open Access (Flächendeckende Umstellung auf Open Access möglich 2015)

- A nugget should be **clearly semantically coherent**.

  If the nugget is syntactically coherent, it is trivial (clarification can be found in nearly all examples).

  If the nugget is syntactically incoherent (but fulfils all other criteria) the following points should be considered:
  - A nugget should be annotated in one colour.
  - The parts of the nugget should be connected using the "∞" symbol.

  Er unterhandelte noch, in der ersten Station, mit den Wirtsleuten, über die Art und Weise, wie er seiner wieder los werden könne: als er schon auf Befehl der Polizei, welche davon Wind bekommen hatte, arretiert und unter einer Bedeckung, er, sein Sohn und Nicolo, so hieß der kranke Knabe, wieder nach Ragusa zurück transportiert ward.
  Example 1.2 Annotation of several nuggets in one sentence (von Kleist 1984)

- A nugget has the **minimal length** of one verb with at least one of its arguments (subject or object).
- A nugget has the **maximum length** of one sentence.
- A nugget does **not contain a title**.
- A nugget consists of **text only**. Tables, figures or other parts of text do not classify as such.

Special cases of nuggets

If a nugget reflects one of the following cases, it is treated separately. The nugget requires fulfilling **all criteria of a nugget** (special attention should be paid to the first one).

If there is not enough information available to treat the special cases as described, the extracts do not fulfil the requirements of a nugget.

- The content of the nugget is not an objective fact.
  - Nuggets consisting of direct or indirect speech, statements, citations or literary text, have to be assigned a source, if the source is not already contained in the nugget.

    Die Leiterin der Kinder-Akademie, Helen Bonzel, legt besonderen Wert darauf, dass in ihrer "Bildungs- und Freizeiteinrichtung für Kinder und Erwachsene" nicht nur Kultur für Kinder, sondern auch Kultur von Kindern Raum geboten wird.
    Example 1.3 Nugget with source already in text (Löffler 2001a)

    Source
    Senator Richard Burr aus Kalifornien stellte trocken fest: "Die Terroristen gehen nicht fort. Amerika ist weiterhin ihr Ziel. Egal, was wir hier im Senat sagen, wir bleiben in ihrem Fadenkreuz."
    Example 1.4 Annotation of the speaker „Senator Richard Burr" fitting his statement (US-Bürger laut Umfrage für NSA-Programme 2015)

    Source
    Der Findling
    […]
    Er unterhandelte noch, in der ersten Station, mit den Wirtsleuten, über die Art und Weise, wie er seiner wieder los werden könne: als er schon auf Befehl der Polizei, welche davon Wind bekommen hatte, arretiert und unter einer Bedeckung, er, sein Sohn und Nicolo, so hieß der kranke Knabe, wieder nach Ragusa zurück transportiert ward.
    Example 1.5 Annotation of several nuggets in a literary text (von Kleist 1984)

    If possible, the annotated source should be directly referring to the statement, e.g. in Example 1.4 not any annotation of "Senator Richard Burr", but the one in direct connection to the statement is chosen. The source can also be a pronoun, which will be resolved in a following step.

- If a potential nugget is an answer to a question without which the nugget would be incomprehensive, it is treated as a **question/answer pair**.
  In the nugget annotation, the answer nugget is marked as usual. If the nugget is a subjective statement, the source has to be added as described in the corresponding paragraph. In this case, the nugget does not have to consist of a verb and a corresponding argument, if the question contains these elements.
  Example 1.6 shows an exemplary annotation of an interview source text, whereas **Fehler! Verweisquelle konnte nicht gefunden werden.** shows an exemplary annotation in a conversation setting.

  Question
  (Taz)   Ist es richtig, neben dem achtjährigen Gymnasium auch das neunjährige anzubieten?
  Source
  Horst Weishaupt   Absolut, solange man nicht grundsätzlich am G8 rüttelt
  Example 1.6 of a nugget that is incomprehensive without the corresponding question (Kramer 2012); Brackets and formatting my author

Source | Question
Sprecher A: Haltet ihr Tierversuche für unnötig?
Source
Sprecher
Sprecher B: Ja, total!
Source
Sprecher
Sprecher C: Nein, finde ich nicht.

Example 1.7 Example of nuggets containing answers of questions by other speakers in a conversation

- The nugget is **not in the indicative mood**.

    If it is an imperative, the nugget can be marked as usual.

    If it is part an interrogative, there are two options:

        If it is part of a question answer/pair, look up how these are treated.

        If there is no answer to the question, the extract cannot be marked as a nugget.

- The potential nugget is not in the language of the final summary.
    o If the extract is or contains a slogan or a special term in another language, it can be annotated. Otherwise it is not considered a nugget.
- The nugget contains listings that are not separated by commas (but with e.g. bullet points).
    o The nugget is annotated as usual; it will be treated differently in Step 5.
- The nugget contains references.
    o If the reference is integrated into the text, so that the nugget would lose or change its meaning without it, it is part of the nugget.

    Da STEWARDs Contrast-Culture-Trainings als 'leistungsfähig, aber sehr aufwendig' einge-schätzt werden, wird nun auch allerorten mit 'Culture-Awareness'-Ansätzen experimen-tiert [...]

    Example 1.8 The reference „STEWARDs" is part of the nugget (Die Entwicklung interkultureller Trainings in Deutschland)

    o If the reference is at the end of a phrase, it most probably is not considered part of the nugget.

    [...] , eignen sich auch für den Einsatz in interkulturellen Lernzusammenhängen in der Schule (Bertelsmann Stiftung 1997 und 1999).

    Example 1.9 The reference „Bertelsmann Stiftung 1997 und 1999" is not part of the nugget (Schoof-Wetzig 2007)

    o If the reference refers to a citation, the rule for citations and indirect speech applies e.g. Example 1.4.
- **Examples** are only part of a nugget, if all referenced circumstances (actions, persons, events …) can be resolved. The nugget has to fulfil all criteria; in this case special attention should be paid to the first one.

Process of multi-document annotation

Texts that are provided on a pre-defined topic built the basis for the annotation.

1. Before annotating, read the current paragraph completely. The main focus here is text comprehension, so please pay attention to words that might seem unfamiliar or to foreign words in general. If necessary, look these up.

2. Try to mark the most important facts only. Pay attention to the compliance of all criteria and the treatment of special cases of nuggets

   In the first step, mark all nuggets that describe the topic well and relate to the following aspects:

   - Important acts/processes must be annotated.
   - If the source text contains persons or causes that play an important role in the information value of the summary, they must be mentioned.

After the annotation of all texts on one topic:

3. **Reread** all nuggets and delete all nuggets without information value.
4. Check all nuggets with regard to the following aspects:
   4.1. Topicality
       Which information do the source texts convey?
   4.2. Accuracy
       Nuggets must contain the most important information of the sources.
   4.3. Objectivity
       Do not use subjective statements without a reference to the source

2. Grouping of nuggets

In this step, redundant nuggets are grouped together. After the annotation of all nuggets in all documents on one topic, nuggets with the same semantic content are put together in one group.

[Wichtig ist, dass die Kinder spüren, dass] wir nicht nur Kultur für Kinder machen, sondern auch die Kultur der Kinder hier Raum hat.

Die Leiterin der Kinder-Akademie, Helen Bonzel, legt besonderen Wert darauf, dass in ihrer "Bildungs- und Freizeiteinrichtung für Kinder und Erwachsene" nicht nur Kultur für Kinder, sondern auch Kultur von Kindern Raum geboten wird.

Example 2.1 Group of nuggets with redundant content (1st nugget taken from Löffler (2001a); 2nd nugget taken from Löffler (2001b))

3. Selection of best nugget

First, choose that nugget, from the cluster of nuggets, which contains the most information, fits the final summary in terms of style or which generally fits the final summary the best. Prefer normal nuggets over special cases. Keep the following in mind:

- **Prefer indirect speech and objective statements** over citations, direct speech and literary text.
- **Prefer nuggets in indicative voice**, rather than imperative or interrogative.
- Prefer enumerations separated by comma over enumerations separated with other signs.
- Prefer general statements over concrete examples.
- Prefer nuggets which do not contain co-references.

4. Co-reference resolution

What is co-reference?

Co-reference resolution represents the relationship between two expressions which refer to the same entity. However, if the co-reference is given within the nugget, it is sufficient to be understandable. In the following example the co-reference between "sie" (she) and "Kanzelerin Merkel" (Chancellor Merkel) is shown.

Nach dem türkischen Militärschlag gegen kurdische Stellungen mahnt [Kanzlerin Merkel]₁ "Verhältnismäßigkeit" an. Im Kampf gegen Terroristen sichert [sie] ₁ der Türkei aber Unterstützung zu.

Example 4.1 Example of a necessary corefernce resolution (Militärschlag gegen Kurden: Merkel mahnt Türkei 2015)

When should co-reference be marked?

In these guidelines, only co-references of entities that are either incomprehensive or ambiguous without their context shall be marked (Paice 1980).

In these guidelines, the focus is on semantics. In this specific step, the focus is on textual equivalence of entities, regardless of grammatical differences like number or case. In Example 4.2 the phrase "der deutschen Mannschaft" (the German team) and "sie" (they) differ in number, but are nevertheless equivalent.

[…], war von dem Spiel [der deutschen Mannschaft]₂ begeistert. [...] Dorthin reisten [sie]₂ dann als Favorit, auch weil sie am 26. Mai 1972 zur Einweihung des Olympiastadions in München die Sowjetunion schon mit 4:1 besiegt hatten.

Example 4.2 Example of co-reference between expressions of different number (Wikipedia 2015)

How is co-reference marked?

The co-reference is marked between the incomprehensive expression in a best nugget and a comprehensive semantically equivalent expression in the source text.

In the process, the closest preceding expression should be chosen. If this is not possible, the closest following expression should be chosen. If no comprehensive expression can be found, the nugget is invalid, as it did not fulfil the requirements[1].

Forms of co-reference

In the following, examples of different forms of co-references are shown. A distinction is made between full and partial co-reference. This distinction and also the subdivision of different co-reference forms will not be reflected in the annotation, however, it helps the understanding and finding of co-references. The list is not exhaustive. In general, co-references can be found between phrases, pronouns, nouns and gerunds.

Moreover, in this step question/answer pairs are treated similarly to co-reference, which is discussed in the last subchapter.

Full co-reference

Full co-reference between two expressions exists if by inserting the comprehensive expression the sense of the nugget is fully preserved. The following examples present classes, where this is the case.

---

[1] In the case another best nugget from the group must be chosen (if there were similar nuggets). The nugget must be deleted.

Pronouns

All pronouns that refer to phrases outside the nugget must be marked with a co-reference[2]. If there are several pronouns in a nugget that refer to one phrase, only one co-reference must be solved, if they are 3rd person pronouns. If possible, the personal pronoun should be marked.

Personal pronouns

Example 4.2 und Example 4. show examples of co-referenced personal pronouns.

> Vor einem Jahr, im Februar 2014, sagte [Putin]7 in einer Sitzung im Kreml: […]. Also liegt somit die Verantwortung bei [ihm]7, und man muss sagen, dass sein Abenteuer gescheitert ist.
>
> **Example 4.3 Co-reference annotation of personal pronouns (Selfanow 2015)**

All phrases that are used as personal pronoun, e.g. "my humble self" being a paraphrase for "me", must also be marked with a co-reference. *Dummy pronouns* (like e.g. the *pleonastic it* in "It rains/snows…") do not have a co-reference.

Possessive pronouns

> Als politischer Seiteneinsteiger vermittelte [Berlusconi]4 das Bild des Machers und Erneuerers. […] [Seine]4 Partei Forza Italia schloss sich zur Parlamentswahl im März 1994 mit Finis Postfaschisten und der sezessionistischen Lega Nord von Umberto Bossi zusammen.
>
> **Example 4.4 Co-reference annotation of possessive pronouns (Rücktritt von Berlusconi)**

Relative pronouns

The co-reference resolution of relative pronouns can be circumvented by including the comprehensive phrase in the nugget in the first step. Nevertheless, if an incomprehensive relative pronoun is in a nugget, it must be annotated with co-reference.

> [Der Beschluss]5, [der]5 an diesem Tag gefällt wurde … .
>
> **Example 4.5 Co-reference annotation of relative pronoun**

Demonstrative pronouns[3]

> Was halten Sie von [Reform XY]6?
> [Diese]6 halte ich für überflüssig!
>
> **Example 4.6 Co-reference annotation of demonstrative pronoun**

Conjunctions

Conjunctions that refer to expressions outside the nuggets must also be marked with a co-reference. This is often the case when conjunctions play the role of subjects or objects.

Explicative Conjunctions

Here, similar to the relative pronouns, the co-reference resolution can be circumvented by more thoughtful annotation in the first step. If this was not done, the phrase must be resoluted.

> [Er ist national bekannt]8. [Das heißt]8, man kennt ihn im ganzen Land.
>
> **Example 4.7 Co-reference annotation of an explicative conjunction**

Causal conjunctions

> [Das Ergebnis der Wahl war nicht eindeutig]9. [Deshalb]9 wurde nochmal abgestimmt.
>
> **Example 4.8 Co-reference annotation of a causal conjunction**

---

[2] When in doubt of whether it is a pronoun, the following list may be of help for German:
http://www.deutschonline.de/Deutsch/Grammatik/Pronomen.htm.
[3] Here , this includes demonstrative determiners that are used as demonstrative pronouns .

Pronominal adverbs

[Methode XY]₁₀ funktioniert folgendermaßen: [...]. Der Vorteil [davon]₁₀ ist [...].

Example 4.9 Co-reference annotation of the pronominal adverb *davon*

Auch da wurde die Frage diskutiert, ob man [wieder zum neunjährigen Gymnasium wechseln]₁₂ sollte. Es spricht nichts [dafür]₁₂.

Example 4.10 Co-reference annotation of the pronominal adverb *dafür (Kramer 2012)*

In the case of implicit pronominal adverbs, the full phrase containing it must be marked:

[Methode XY]₁₁ funktioniert folgendermaßen: [...]. [Der Vorteil]₁₁ ist [...].

Example 4.11 Co-reference annotation of an implicit pronominal adverb

Ambiguous Expressions

Especially in the context of several documents, the usage of some expressions may become ambiguous.

Hypernyms

The same hypernym may refer to different hyponyms in different paragraphs or documents. To keep the unambiguity of the phrase, the co-reference must be resolved.

Der [Culture-Awareness-Ansatz]₁₂ ist ein Beispiel
[...]
Die Stärken dieses [Ansatzes]₁₂ liegen vor allem in der universellen Anwendbarkeit
[...]
Der [Contrast-Culture-Ansatz]₁₃ geht davon aus, dass
[...]
Auch für kulturübergreifende Schulungsmaßnahmen ist der [Ansatz]₁₃ nur bedingt geeignet.

Example 4.12 Co-reference annotation of the hypernym „Ansatz" (approach) in multiple contexts („Interkulturelles Lernen" 2015)Intentions, events and actions

Some phrases only make sense with the preceding phrase. In this case, the unambiguity can be preserved through co-references.

[Das Hilfspaket für Griechenland wurde beschlossen]₁₃. [Der Beschluss]₁₃ wurde von vielen als schlecht empfunden.

Example 4.13 Co-reference annotation of an event

Time expressions

If nuggets use ambiguous time expressions such as e.g. today, yesterday, last year (in contrast to September 2001, 10/03/2015, …), a co-reference to an unambiguous time expression must be made .

Am [01.10.2014]₁₅ gingen sie nach China. [Zwei Tage darauf]₁₅ reisten sie wieder zurück.

Example 4.14 Co-reference annotation of time expression

Partial co-reference

A partial co-reference occurs when the content or incomprehensive and ambiguous expressions refers to several phrases.

"Sie" (they) is co-referent to *Hollande* und *Merkel* in Example 4..

[Hollande]₃_part und [Merkel]₃_part hatten zuvor mitgeteilt, [sie]₃ würden zu einer neuen Friedensinitiative nach Kiew und Moskau aufbrechen.

Example 4.15 Co-reference annotation of "sie" (they)

Question-answer pairs

In the case of question answer pairs, the question is marked similarly to a co-reference of the answer. The fact that it is not a classical co-reference is reflected in the annotation by adding a "Question"-tag at the beginning of the question reference. Example 1.6 is annotated in the following way:

Horst Weishaupt   Absolut[Question: Ist es richtig, neben dem achtjährigen Gymnasium auch das neunjährige anzubieten?], solange man nicht grundsätzlich am G8 rüttelt

Example 4.16 Co-reference annotation of Example 1.6 (Kramer 2012); Brackets and formatting my author

5. Reformulation into sentences

The best nuggets are reformulated to grammatically correct sentences. The following points should be considered:

- The information value of the nugget shall remain unaltered.
  - Adding information that is not the nugget is not allowed.
    Hence, in Example 5.1 below „sie" (she) cannot be replaced with "*Bundeskanzlerin Merkel" (State chancellor Merkel)*, *Angela Merkel* or "*die amtierende Bundeskanzlerin* " (the current chancellor), because this information is not in the source text. Rather, it is replaced with "Kanzlerin Merkel", because that is the co-references relation in the source text.

Co-references are solved by replacing the ambiguous or incomprehensive with clear phrases.

Example 4.1 is reformulated in the following way:

Nach dem türkischen Militärschlag gegen kurdische Stellungen mahnt **[Kanzlerin Merkel]₁** "Verhältnismäßigkeit" an.
Im Kampf gegen Terroristen sichert **[sie] ₁** der Türkei aber Unterstützung zu.
→ Im Kampf gegen die Terroristen sichert Kanzlerin Merkel der Türkei aber Unterstützung zu.
Example 5.1Reformulation of a nugget to a sentence

After insertion, changes to make the sentence grammatically correct should be performed.

Example 4.2 is reformulated in the following way:

[…], war von dem Spiel **[der deutschen Mannschaft]₂** begeistert. […] Dorthin reisten **[sie]₂** dann als Favorit, auch weil sie am 26. Mai 1972 zur Einweihung des Olympiastadions in München die Sowjetunion schon mit 4:1 besiegt hatten.
→ Dorthin reiste die deutsche Mannschaft dann als Favorit, auch weil sie am 26. Mai 1972 zur Einweihung des Olympiastadions in München die Sowjetunion schon mit 4:1 besiegt hatte.

Example 5.2 Grammatical reformulation of nugget with co-reference

The special cases described in the 1st step should be treated in the following way:

- The sources of the subjective statements must be integrated.
  - Direct speech must be reformulated to indirect speech, applying adjustments such as the subjunctive[4].

Source
König: […] Kinder grundsätzlich anders lernen als Erwachsene.

→ Museumspädagogin König sagt, dass Kinder grundsätzlich anders als Erwachsene lernen würden.
Example 5.3 Reformulation of nugget with direct speech to indirect speech

- Nuggets from literary text must be reformulated, so that they contain the name of the title.

Hence, the nugget in Example 1.5 is reformulated in the following way:

Ursprungswerk
Der Findling
[…]

---

[4] For this, a "Konjunktor" for German such as http://konjugator.reverso.net/konjugation-deutsch.html may be used.

Er [Piachi]unterhandelte noch, in der ersten Station, mit den Wirtsleuten, über die Art und Weise, wie er seiner wieder los werden könne: als er [Piachi] schon auf Befehl der Polizei, welche davon Wind bekommen hatte, arretiert und unter einer Bedeckung, er, sein Sohn und Nicolo, so hieß der kranke Knabe, wieder nach Ragusa zurück transportiert ward

➔ In „Der Findling" unterhandelt er [Piachi] noch mit den Wirtsleuten.
➔ In „Der Findling", wurde er [Piachi]auf Befehl der Polizei arretiert und wieder nach Ragusa transportiert.

Example 5.4 Reformulation of two nuggets (from Example 1.5) from a literary source text

- Nuggets from question/answer pairs (potentially with a speaker) must be reformulated to a grammatically correct sentence. The question shall be reformulated to a statement, containing the answer (Bar-Haim et al. 2006; Giampiccolo et al. 2007). If applicable, the speaker shall be added as discussed in the previous step, using the subjunctive in the statement.

Example 1.6 is reformulated in the following way:

	Frage
(Taz)	Ist es richtig, neben dem achtjährigen Gymnasium auch das neunjährige anzubieten?
Sprecher	
Horst Weishaupt	Absolut, solange man nicht grundsätzlich am G8 rüttelt

➔Horst Weishaupt sagt, dass es absolut richtig sei, neben dem achtjährigen Gymnasium auch das neunjährige anzubieten, solange man nicht grundsätzlich am G8 rüttle.

Example 5.5 Example einer Umformung eines Nuggets, das aus einem Frage-Antwort Paar (aus Example 1.6) besteht

- Nuggets in the imperative should be reformulated to the indicative.

  Bürger, denkt nach!
  ➔ Die Bürger sollen nachdenken.
  Example 5.6 Reformulation of a nugget in the imperative

- Listings, that are not separated by a comma should be treated in the following way:
  o If the listing contains 4 items at most, the nugget should be reformulated to a sentence with commas.

Kommunikation bedeutet:
  o Verständigung
  o Mitteilung / Information
  o Interesse
  o Austausch

➔

Kommunikation bedeutet Verständigung, Mitteilung / Information, Interesse und Austausch.

Example 5.7 Reformulation of nugget with at most 4 items; Left: nugget, Right: reformulated sentence (Gottschalk)

- If the listing contains more than 4 items, it should be reformulated with bulletpoints (regardless of the listing form in the nugget)

Die Abfolge der Arbeitschritte oder Prozeduren der Critical Incidents Methode können folgendermaßen dargestellt werden:
1. Auswahl des Untersuchungszieles
2. Bestimmung des Untersuchungskontexts
3. Bestimmung der Beteiligten
4. Entwicklung eines Beobachtungskatalogs
5. Entwicklung eines Fragenkatalogs
6. Erhebungsphase
7. Auswertungsphase
8. Interpretation der Ergebnisse, Schlussfolgerungen

Die Abfolge der Arbeitschritte oder Prozeduren der Critical Incidents Methode können folgendermaßen dargestellt werden:
- Auswahl des Untersuchungszieles
- Bestimmung des Untersuchungskontexts
- Bestimmung der Beteiligten
- Entwicklung eines Beobachtungskatalogs
- Entwicklung eines Fragenkatalogs
- Erhebungsphase
- Auswertungsphase
- Interpretation der Ergebnisse, Schlussfolgerungen

Example 5.8 Reformulation of nugget with more than 4 items; Left: nugget, Right: reformulated sentence **(Göbel 2003)**

6. Structuring

To facilitate the final process of summary writing, the sentences are structured.

This step consists of two substeps. In the first substep, sentences that could be in the introduction are selected. In the second step groups freely chosen by the annotator to structure the content further are built.

Selecting sentences for the introduction

The introduction should contain the most important aspects of the overall summary. The following aspects may support the selection of sentences for the introduction.

- Nuggets that contain descriptions that are relevant for the understanding of the topic should be in the introduction. Descriptions and definitions in the introduction facilitate the flow of reading.

  In Example 1.1 the topic *Open Access* is defined. Hence, the therein shown nugget („… Open Access im Sinne dieser Resolution bezeichnet den freien, für die Nutzer im Regelfall kostenlosen Zugang zu mit öffentlichen Mitteln produziertem Wissen…" (Open access in the sense of this resolution describes the open, and as a rule cost-free, access to publically funded knowledge…")) should be grouped as *Introduction*.

- Nuggets, containing Named Entities (NEs)[5], meaning names, should be treated in the following way:

  Informative descriptions of NEs should be grouped under *Introduction*.

  In Example 6.1 the first sentence should be grouped under Introduction, because it contains the function "*ukrainischer Präsident*" (Ukrainian President) of the NE "*Petro Poroschenko*" (Petro Poroshenko):

  Petro Poroschenko versprach Frieden und Wirtschaftsreformen.
  [...]   Introduction
  Es ist erst ein paar Wochen her, da hatte der ukrainische Präsident Petro Poroschenko seinen großen Auftritt auf internationalem Parkett. Er hielt in der Schweiz eine viel beachtete Rede, im Rahmen des Weltwirtschaftsforums in Davos.
  [...]
  Poroschenko hatte seinen Bürgern einen schnellen Sieg versprochen.
  Example 6.1 Possible grouping of informative description of person as Introduction (Bidder 2015)

  In Example 6.2 the second sentence contains the function "*Burkina Fasos Hauptstadt*" (capital of Burkina Faso) of the NE "*Ougadougou*":

  In Ouagadougou blieb den Menschenmengen nur der Zugang zum abgeschotteten Präsidentenpalast mit Scharfschützen auf dem Dach verwehrt.
  [...]   Introduction
  Die aufsässigen Jugendlichen, die am 30. Oktober Burkina Fasos Hauptstadt Ouagadougou verwüsteten und am Folgetag Präsident Blaise Compaoré stürzten, wollten wohl keinen 72-Jährigen als Nachfolger.
  Example 6.2 Possible grouping of informative description of location as Introduction (Sentence 1 taken from Johnson (2014b); Sentence 2 taken from Johnson (2014a))

---

[5] Named Entity is a term for names of persons, locations, organizations and other groups For more detailed information see Benikova et al. 2014.

Subtopic clustering

To perform subtopic clustering, sentences that refer to a subtopic of the topic are grouped together.

In Example 6.3 such a subtopic clustering on the topic *Number of school years in Germany* is shown. The titles of the individual clusters are bold and underlined. The sentences in the clusters are separated by horizontal lines.

<u>Unterthemengruppe *G8/G9 in Bayern*</u>

An insgesamt 370 Gymnasien in Bayern wurden seit Herbst 2011 Lehrkräfte zum G8-Lehrplan befragt

Der Lehrplan des Gymnasiums [in Bayern] erfährt in fachlicher Hinsicht insgesamt hohe Zustimmungswerte.

Die Regierungsfraktion des Bayerischen Landtags hat eine Resolution zur Weiterentwicklung des bayerischen Gymnasiums beschlossen

Wir [in Bayern] sind die ersten, die in der Lage sind, den Gegensatz von G8 und G9 aufzuheben

<u>Unterthemengruppe *G8/G9 in Hessen*</u>

Die hessischen Regierungsparteien CDU und BÜNDNIS 90/DIE GRÜNEN haben in ihrem Koalitionsvertrag festgelegt, dass sie die Wahlfreiheit der Schulen zwischen G8 und G9 an den kooperativen Gesamtschulen und den Gymnasien weiter stärken wollen.

Hessen hat die Wahlfreiheit für Schulen zwischen G8 und G9 eingeführt

Die Gymnasien und kooperativen Gesamtschulen [in Hessen], die ab dem Schuljahr 2014/2015 von G8 zu G9 wechseln wollen, können die jeweils bestehenden Jahrgänge 5, 6 und 7 in diesen Wechsel

An Schulen [in Hessen], die bereits zu Beginn des Schuljahres 2013/2014 zu G9 gewechselt sind, wird für die bestehenden Jahrgänge 6 und 7 ein Wechsel ebenfalls ermöglicht.

Example 6.3 Subtopic clustering on the topic of „Number of School Years in Germany"according to states in Germany

7. Summary creation

In this step the final summaries are created, using the sentence structure that was discussed in the previous step.

The following should be considered:

- No additional information must be added.
- The text should be coherent. The following points may support this:
  - The so called sequence of tenses should be followed, meaning that the described events/actions should be described using the appropriate tenses. The same event cannot be described in both the present and the past tense in the same summary.
  - Connectives and conjunctives should be used.
  - All co-references must be resolved[6].
- It must be a grammatically correct text.
- The summary should be understandable without the source texts.
- The structure of the previous step does not have to be used (But is probably useful).

If the summary is considerably longer than 500 words, a second summary with at most 500 words (preferably less) has to be produced by cutting out less important sentences or phrases. If necessary, the left sentences or parts of sentences have to be reformulated to form a grammatically correct text.

---

[6] It is allowed not to resolve co-references if all entities can be derived from the context of the summary.

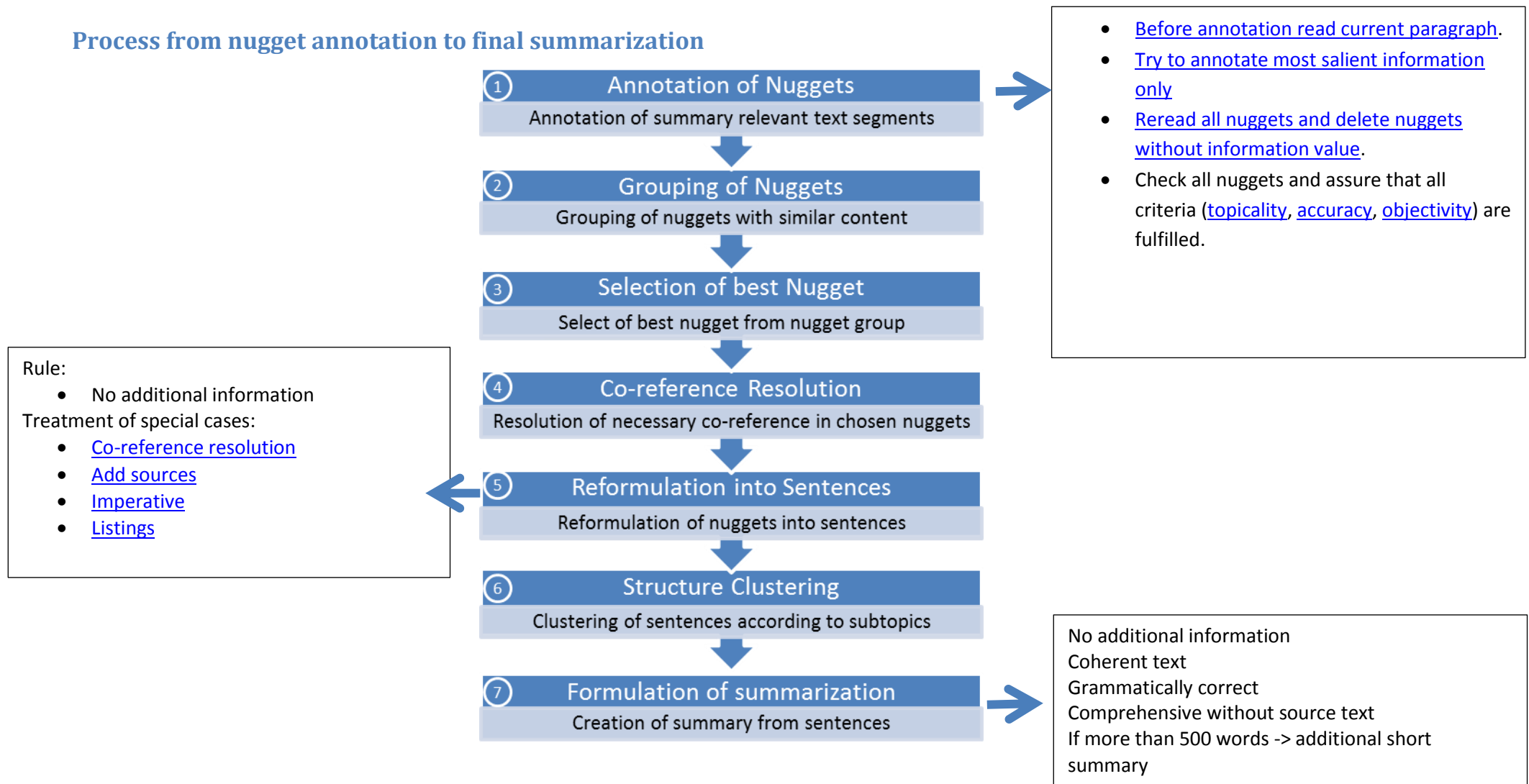# Short manual for annotation of nuggets and summary creation

Nuggets

Nugget criteria
- [Answer to topic question](#)
- [Clearly cohesive](#)
- [Min. length](#)
- [Max. length](#)
- [No titles](#)
- [Text only](#)

Special cases
- [Not objective fact](#)
- [Not indicative](#)
- [Foreign language](#)
- [Listings](#)
- [References](#)
- [Examples](#)

Process from nugget annotation to final summarization

- Before annotation read current paragraph.
- Try to annotate most salient information only
- Reread all nuggets and delete nuggets without information value.
- Check all nuggets and assure that all criteria (topicality, accuracy, objectivity) are fulfilled.

① **Annotation of Nuggets**

Annotation of summary relevant text segments

② **Grouping of Nuggets**

Grouping of nuggets with similar content

③ **Selection of best Nugget**

Select of best nugget from nugget group

④ **Co-reference Resolution**

Resolution of necessary co-reference in chosen nuggets

Rule:
- No additional information

Treatment of special cases:
- Co-reference resolution
- Add sources
- Imperative
- Listings

⑤ **Reformulation into Sentences**

Reformulation of nuggets into sentences

⑥ **Structure Clustering**

Clustering of sentences according to subtopics

⑦ **Formulation of summarization**

Creation of summary from sentences

No additional information
Coherent text
Grammatically correct
Comprehensive without source text
If more than 500 words -> additional short summary

Publication bibliography

Die Entwicklung interkultureller Trainings in Deutschland. In *Arbeitspapiere des IFIM*. Available online at http://www.ifim.de/reports/entwicklung.pdf.

Rücktritt von Berlusconi. Kabale, Triebe - und ein Halleluja. Available online at http://www.sueddeutsche.de/politik/silvio-berlusconi-der-kampf-des-cavaliere-1.1035501, checked on 9/16/2015.

Open Access – Chancen für den Zugang zum Wissen für alle (2007). In *Deutsche UNESCO-Kommission e.V.* Available online at https://www.unesco.de/infothek/dokumente/resolutionen-duk/reshv67-3.html.

Flächendeckende Umstellung auf Open Access möglich (2015). In *Max-Planck-Gesellschaft*. Available online at http://www.mpg.de/9201460/flaechendeckende-umstellung-auf-open-access-moeglich.

„Interkulturelles Lernen" (2015). In *Wikipedia, Die freie Enzyklopädie*. Available online at https://de.wikipedia.org/w/index.php?title=Interkulturelles_Lernen&oldid=143104464.

Militärschlag gegen Kurden: Merkel mahnt Türkei (2015). In *Spiegel Online*. Available online at http://www.spiegel.de/politik/ausland/merkel-appelliert-an-tuerkei-nach-militaerschlag-gegen-pkk-a-1045411.html.

US-Bürger laut Umfrage für NSA-Programme (2015). Available online at http://www.tagesschau.de/ausland/nsa-senat-109.html.

Bar-Haim, Roy; Dagan, Ido; Dolan, Bill; Ferro, Lisa; Giampiccolo, Danilo; Magnini, Bernardo; Szpektor, Idan (2006): The second pascal recognising textual entailment challenge. In : Proceedings of the second PASCAL challenges workshop on recognising textual entailment, vol. 6, p. 6.

Benikova, Darina; Biemann, Chris; Reznicek, Marc (2014): Nosta-d named entity annotation for german: Guidelines and dataset. In : Proc. LREC.

Bidder, Benjamin (2015): Ukrainischer Staatschef Poroschenko: Der Kriegspräsident ist angezählt. In *Spiegel Online*. Available online at http://www.spiegel.de/politik/ausland/ukraine-krise-kritik-an-praesident-poroschenko-waechst-a-1019394.html.

Buhre, Jakob; Baumbach, Marie von (2013): Die politische Klasse hat unverdienter Maßen ein grundnegatives Image.

Giampiccolo, Danilo; Magnini, Bernardo; Dagan, Ido; Dolan, Bill (2007): The third pascal recognizing textual entailment challenge. In : Proceedings of the ACL-PASCAL workshop on textual entailment and paraphrasing. Association for Computational Linguistics, pp. 1–9.

Gottschalk, Bernd: Grundlagen der Kommunikation. Available online at http://www.dpsg-ingelfingen.de/images_Library/Leiterwochenende%202005/Dokumente/Grundlagen%20der%20Kommunikation.pdf.

Johnson, Dominic (2014a): Burkina Fasos Übergangspräsident. Der alte Kafando für den Neuanfang. Available online at http://www.taz.de/Burkina-Fasos-%C3%9Cbergangspr%C3%A4sident/!5028438/.

Johnson, Dominic (2014b): Volksaufstand in Burkina Faso. Kraftprobe in Ouagadougou. In *taz.de*. Available online at http://www.taz.de/!5029776/.

Kramer, Bernd (2012): Bildungsforscher über Abiturdauer. Es spricht nichts gegen das G8. Available online at http://www.taz.de/!5087586/.

Lin, Jimmy; Demner-Fushman, Dina (2006): Will pyramids built of nuggets topple over? In : Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Association for Computational Linguistics, pp. 383–390.

Löffler, Udo (2001a): Fitnesscenter der Sinne. Kindermuseen als Bildungsorte auf dem Vormarsch. In *Bildung+Innovation*. Available online at http://www.bildungsserver.de/innovationsportal/bildungplus.html?artid=123.

Löffler, Udo (2001b): "Kindermuseen sind keine überdachten Abenteuerspielplätze". Museumspädagogin König über den Bildungsort Kindermuseum. In *Bildung+Innovation*. Available online at http://www.bildungsserver.de/innovationsportal/bildungplus.html?artid=54.

Nenkova, Ani; Passonneau, Rebecca; McKeown, Kathleen (2007): The pyramid method: Incorporating human content selection variation in summarization evaluation. In *ACM Transactions on Speech and Language Processing (TSLP)* 4 (2), p. 4.

Padtberg-Kruse, Carola (2013): Übertritt aufs Gymnasium: Lehrer wählen besser aus. In *Spiegel Online*. Available online at http://www.spiegel.de/schulspiegel/gymnasialempfehlung-elternwille-macht-schulsystem-noch-ungerechter-a-938615.html.

Paice, Chris D. (1980): The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases. In : Proceedings of the 3rd annual ACM conference on Research and development in information retrieval. Butterworth & Co, pp. 172–191.

Schoof-Wetzig, Dieter (2007): Training interkultureller Kompetenz. In *Interkulturelle Bildung*. Available online at http://www.nibis.de/nli1/ikb/iktraining.html.

Selfanow, Konstantin (2015): Professor Subow: Die Generalität des KGB ist mit Putin unzufrieden. In *Euromaidan Press*. Available online at http://de.euromaidanpress.com/2015/05/31/professor-subow-die-generalitaet-des-kgb-ist-mit-putin-unzufrieden/, checked on 9/16/2015.

von Kleist, Heinrich (1984): Der Findling. Stuttgart: Philipp Reclam jun. GmbH & Co.

Wikipedia (2015): Deutsche Fußballnationalmannschaft -- Wikipedia, Die freie Enzyklopädie. Available online at https://de.wikipedia.org/w/index.php?title=Deutsche_Fu%C3%9Fballnationalmannschaft&oldid=145899425.