Creating Instructional Content for Augmented Reality based on Controlled Natural Language Concepts

Steffen Mader, Bodo Urban
Fraunhofer Institute for Computer Graphics, Rostock, Germany

ABSTRACT

In this paper we present an approach to support the content generation for Augmented Reality (AR)-based technical manuals by modeling and interpreting technical instructions in a semantical way. We propose the separation of the meaning of a technical instruction from its actual representation and accomplish this by designing a semantic instruction model on the base of a controlled natural language. The controlled language serves here as a high-level specification model for technical activities. Its controlled lexicon and grammar eliminate the ambiguity of meaning and reduce the complexity of language constructs known from full natural languages. This enables automated processing of the instructions. Complemented with references to the geometry of assemblies, parts, tools and other geometric features, the resulting formalized instructions can be easily transformed to both textual and graphical representations. In the demonstration of our results we will focus on the generation of scenes for AR-based maintenance support.

KEYWORDS: Augmented Reality, Authoring, Controlled Natural Languages.

1 INTRODUCTION

Augmented Reality has successfully proven its applicability as a supportive technology in service and maintenance scenarios. One of the major factors which limit the forthcoming/acceptance of this technology, however, is the way technical documentation is produced today. The traditional technical documentation process usually generates illustrated textual, page-oriented manuals that are commonly published on paper or CD-ROM. At the level of single instructions, those materials usually combine a textual instruction part with a 2-d graphical illustration complementing the text. As such they are not able to provide a suitable base for the generation of AR-content. Hence, an additional authoring process is needed to generate the AR-content, which boosts the overall costs of the technology.

Although there already exist a number of solutions that optimize the additional authoring process for certain aspects, a reformation of the technical documentation process regarding the production of interactive media is inevitable. First modern approaches emerge that place the authoring of interactive media into the center of the documentation process, generating traditional media only as a side-product [3].

Fraunhofer Institute for Computer Graphics
Joachim-Jungius-Str. 11
18059 Rostock, Germany
{steffen.mader | bodo.urban}(at)igd-r.fraunhofer.de

20th International Conference on Artificial Reality and Telexistence (ICAT2010), 1-3 December 2010, Adelaide, Australia
ISBN: 978-4-904490-03-7 C3450 ©2010 VRSJ

Whereas this might solve the problem of the duplicate processes, it still does not solve the problem of the duplicate semantics raised by the parallel management and storage of different representations of the same semantic fact (e.g. instructional text + illustration). With the increasing importance of interactive media in technical documentation, the number of representations for the same technical fact will further rise and lead to a multiplication of the semantic redundancy.

In our contribution we propose the design of a semantical instruction model that provides a separation of the semantic meaning within technical instructions from its actual representations. The envisioned model will be able to describe technical instructions on a level that allows the derivation of diverse forms of medial representation from a single source.

In the following, the paper is structured as follows: In section 2 we give a brief definition of controlled natural languages, and shortly introduces the language used within our work. In section 3 we will give an overview on the concepts and approaches that have been used so far to support the content creation of instructional AR-scenes and on the use of CNL in related fields. Following this, we will introduce the concepts making up our approach and give a short introduction to our formalized instruction model, as well as we will briefly describe the transformation process used to produce textual and graphical representations of the formalized instructions and present first results of the generated content. Finally, we will conclude and summarize the advantages of our approach and give an outlook on our further work in this field.

2 CONTROLLED NATURAL LANGUAGES

2.1 General Definition

Following the definition provided by Schwitter and Tilbrook [11], controlled natural languages (CNL) are limited subsets of natural languages with explicit restrictions on the grammar, lexicon, and style. These restrictions usually have the form of writing rules and help to reduce (or even exclude) ambiguity and to cut down the complexity of full natural language.

According to their purpose, controlled natural languages subdivide into two categories: human-oriented and machine-oriented controlled natural languages.

While human-oriented CNL are intended to improve the text comprehension for human readers, machine-oriented CNL are designed in order to make text machine processable.

2.2 Simplified Technical English

Since our approach aims at the area of technical documentation we focus our work on the use of Simplified Technical English (ASD-STE100 [1]). ASD-STE100 is an international standard that specifies a controlled natural language for the purpose of technical documentation. It originates from an aerospace and defense standard, and is strictly defined by a set of 57 writing rules and a limited vocabulary of approx. 900 approved words with well defined and unique meanings. Its vocabulary is extendable through the definition of domain-specific dictionaries

3 Related Work

3.1 AR-Authoring

Much work has been conducted in order to optimize the authoring process for VR/AR-based technical documentation and to reduce the costs of the additional authoring for the 3-d scenes. The optimization strategies of the common approaches mainly rely on the principles of task-oriented interaction design, component re-use and data integration.

Supporting the author with intuitive forms of interaction is the most common principle in this field. Several authoring tools for AR-applications provide task-tailored 3-d scene manipulation and/or visual programming paradigms for the design of work- and data flows as well as for the definition of relations between the involved entities [3][4][5][8][13].

Component re-use is done by the definition of 3-d templates for frequently occurring action patterns. Instances of these 3-d templates can be parameterized to fit the specific use case. Complex templates might comprise complete 3-d scenes that involve highlighting or animation of parts and tools [3][5][15], whereas simpler templates might merely form a construction kit providing the geometric primitives of common symbolic augmentations and tools.

Data integration is an aspect that gained more attention during the recent years. Instead of keeping technical documentation data separated from the product data management (PDM), the data are stored within the product model in the form of meta-data. This enables the integration of technical documents into the change management process and supports component-wise re-use of documentation [13][14].

Another concept for the support of authoring in the fields of virtual and augmented reality is the use of high-level description models in order to generate scenes and screenplay. Such models allow the formalization of content structures and behavior. Moreover, they support modularization and partition of both content and authoring aspects, and thereby directly advocate the concept of re-use. The Augmented Presentation and Interaction Language (APRIL) [7] is one of the implementations of this concept. With APRIL, authors may describe states of a complex scenario and map interaction to transitions between those states. The state transitions may even describe simple animations of scene parts. However, APRIL focuses on narrative scenarios rather than instructive ones and relies on prior modeled scene parts. It does not provide functionality to model or generate them.

3.2 Controlled Language for Knowledge Representation

There are approaches to the use of controlled languages for knowledge representation in several related fields.

Schwitter and Tilbrook proposed to use CNL for the semantic annotation of websites [12] with machine-processable information. For this purpose they designed PENG (Processable ENGlish), a machine-oriented controlled natural language for summarizing individual web pages of a website and to express domain-specific ontological knowledge about that website in an unambiguous subset of English. Further on, the annotations are processed to first-order logic constructs and reasoning models that can be used for consistency and informativity checking as well as for question answering.

Semantics of Business Vocabulary and Business Rules (SBVR) [10] is an adopted standard of the Object Management Group (OMG) intended to be the basis for formal and detailed natural language declarative description of a complex entity, such as a business. SBVR is intended to formalize complex compliance rules, such as operational rules for an enterprise, security policy, standard compliance, or regulatory compliance rules. Its formal vocabularies and rules can be interpreted and used by computer systems.

Similar to our approach, Martignoni, Smart [9] and Kress-Gazit et.al. [6] propose controlled natural languages to formulate high-level task descriptions. The focus of their work, however, lies on the motion control of mobile robots rather than in the single source production of technical documentation materials.

All the above approaches have in common, that they use languages that are specifically designed for their original purpose. Our approach however, tries to utilize an existing, well established language specification (ASD-STE100 [1]) within its original, primary domain except for additional purpose.

4 Approach

4.1 Concepts

As stated above, we identified two major shortcomings in the traditional documentation process and the current practice of authoring content for AR-based technical documentation: the duplicate authoring process and the multiplication of semantics in texts and illustrations.

In order to overcome these shortcomings the key concept of our approach is it to model and handle only one single, high-level representation of an instruction, which then serves as the source for media production. Our approach, therefore, adds another abstraction level on top of those multiple representations in order to be able to:

- eliminate the semantic redundancy induced by the different media,
- produce all needed representations (e.g. textual, schematic, graphical) from a single source of information on the base of a single authoring process.

This poses certain requirements on such a single abstract information source. While this source should be able to provide the basis to produce a well-formed natural language text for display in page-oriented documentation such as paper or pdf-documents, it should also contain all the information necessary to reflect the meaning of the instruction in a symbolic way and to register it within the correct spatial context of the target plant or product in order to generate illustrations or even animated 3-d scenes.

Our approach addresses these requirements based on a few simple concepts:

1. *Elimination of natural language ambiguity by use of a controlled natural language:*

We eliminate the ambiguity of natural language by using the restricted vocabulary of a controlled natural language, esp. ASD-STE 100. The STE-specification assigns a strict usage (part of speech) and a single, unique meaning to every term in the dictionary, thus fixing the semantics of the terms, and providing a suitable lexical base for our approach allowing the automatic generation of appropriate representations for actions (e.g. remove, turn, examine), their modes (e.g. manner, time, frequency, duration) and physical properties of the involved parts (e.g. colors, shapes, materials etc.).

2. *Further reducing the complexity of natural language by formalizing the semantics and grammar of instructive sentences:*

ASD-STE 100 itself already defines several writing rules that serve the purpose of restricting the complexity of language constructs. This includes: demanding sentences to be written only in an active voice, restricting sentences to express only one single factum, or limiting the use of tenses to only simple present, simple past and simple future tenses. In addition to these rules, we further reduce the complexity of sentences by casting the involved parts of speech into a fixed scheme of grammatical (e.g. predicative phrases, noun phrases) and semantic (e.g. activity, target, tool) roles within the instructive sentence. Since technical authoring is a creative process (rather than analytical) we can easily assign these roles during authoring.

3. *Provision of the spatial/geometrical context of the activity by providing references into the geometrical model of the Product or environment:*

This involves the establishment and preservation of references to the involved parts, tools, supplies. Preferably, these references are obtained by data structures exported from the product data management (PDM) system. Beside the spatial positions of parts and prominent features (e.g. bores, slots, handles) which are necessary to correctly position the generated symbolic augmentations, these data structures should also provide information about part numbers, part names, properties, connection types etc. Our implementation currently uses a proprietary format that originates from our work within the AVILUSplus [2] project. This specialized format contains all the necessary information by design, but research is done on other existing exchange formats in order to provide an open, standardized solution.

Following these three principal concepts allows us the semantically unambiguous definition of technical instructions, the involved entities and their respective spatial allocations. The resulting definition is independent of a later concrete (textual | visual | aural) representation. Thus, it is able to provide a clear separation between content and medial representation.

4.2 Implementation

In this section we give a short overview on the implemented specification. For the reason of clarity and available space we will only present a small part of the design.

Due to practical reasons we decided to use XML as the implementation base of the `<semantx>`-scheme. XML is a widely used standard for describing structural data, and new dialects can be standardized with DTD’s and Schemas, which are also widely used and understood. XML allows the incorporation of any other ASCII-based file formats (by simply including the data inside an elements CDATA section) or other XML-based formats (through the concept of namespaces) into documents. This allows integration of our framework into various other XML formats existing for technical documentation. Finally, a whealth of tools are available that operate on XML data, such as parsers, validators or XSLT, a technology to transform XML data into other document formats. As we will show further below, we use XSLT to transform the formalized instructions into textual and graphical representations. With the choice for XML as the base technology for our content format, the next step was to design the structure of the specification model.

The overall structure of the description model is made-up by a hierarchy of grammatical and semantical categories. Since the grammatical categories often coincide with semantical categories, their strict separation is not always possible. Moreover, the explicit modeling of certain grammatical categories alleviates a later translation into proper textual representations without posing negative side-effects onto other forms of representation.

On the top-level of the structure are the semantical categories. They define the instruction in terms of the action to accomplish and the participating entities.

```
<instruction>
   <action_phrase/>
   <target_phrase/>
   <tool/>
</instruction>
```

An action can be further specified by an action verb and one or more modifiers that provide certain modes or parameters. The range of available action verbs and their word semantics are strictly defined by the ASD-STE100 specification. This also accounts for modifying adverbials and prepositions.

```
<action_phrase>
   <action_verb action=”remove”/>
   <mode>
      <mode_manner manner=”cautiously”/>
   </mode>
</action_phrase>
```

The target phrase is actually a noun phrase, where the noun takes the semantic role of the instruction’s target. Targets and nouns may be further qualified within their respective phrases, where the range of available adjectives, again, is defined by the ASD-STE100 specification. Targets and nouns contain references to entities (assemblies, parts, features) from the geometry or product model.

```
<target_phrase>
   <target>
      <part part_id=”DMC-AE-A-04-…-00”/>
   </target>
   <qualifier>
      <qual_colour colour=”red”/>
   </qualifier>
</target_phrase>
```

In order to resolve those references, a linkage to the referenced model must exist. This can be done in either two ways: by integration of the semantx-scheme into a subordinate scheme providing the referenced entities, or by access to a separate model containing these entities via shared key structures.

The transformation of formalized instructions into textual and graphical representations is done by application of XSL-Transformations (XSLT). XSL is a template-based language for transforming XML documents into other, ASCII-based document types. One or multiple input files can be processed in a non-linear fashion, generating arbitrary numbers of output files. XSL code that is used to generate HTML output from XML documents is generally referred to as stylesheets, where the term XSLT refers to XSL-code that is used to transform and aggregate collections of XML documents into other XML documents.

Our implementation so far provides transformations into two different target formats. One of the transformations generates a plain text string from the formalized instruction, representing the instruction in Simplified Technical English conforming to the specification ASD-STE100 [1]. The second transformation produces an Augmented Reality scene in X3D format, representing the technical action within its spatial context in a symbolic way.

Since names of parts, spatial context and geometry have to be resolved from references to separate or superordinate content, i.e. product data, the transformations also depend on the format and semantics of that content. Our implementations so far relate to a proprietary format that originates from our work within the AVILUSplus [2] project. However, other implementations that refer to standardized or widely adopted source formats will be subject of our further work.

5 Results

Let us now demonstrate our results with a small example: Assume, we want to give out the instruction to “remove the screws that hold the hard drive using a Philips screw driver size PH2”. Listing 1 shows a snippet of the formalized instruction, where we omitted some minor details for clarity reasons.

```
<instruction>
   <action_phrase>
      <action>
         <verb_remove/>
      </action>
   </action_phrase>
   <tool tool_id="t_001" size="PH2"/>
   <target_phrase>
      <target>
         <part part_id="p_020"/>
      </target>
      <target>
         <part part_id="p_021"/>
      </target>
      <qualifier>
         <qual_rel rel_pron="that">
            <rel_clause>
               <verb_phrase>
                  <verb>
                     <verb_hold/>
                  </verb>
               <verb_phrase>
               <noun_phrase>
                  <noun>
                     <part part_id="p_010"/>
                  </noun>
               </noun_phrase>
            </rel_clause>
         </qual_rel>
      </qualifier>
   </target_phrase>
</instruction>
```

Listing 1. Example code for a complete instruction.

Transforming this instruction to plain text gives us:

Remove the screws that hold the hard disk drive using a Philips screwdriver size PH2!

as expected, with the part names “screws” and “hard disk drive” resolved from the product data. This kind of representation may be used to produce text-based documentation, to generate audible output via text-to-speech or to provide explanatory text in addition to graphical representations.

In order to produce 3-d representations of the instruction, the product data referenced by the formal instruction must provide the geometry of the parts, important geometric features and properties as well as describe their spatial configuration. In our example the product data provide the part geometry in form of an X3D component file for each part, a transformation hierarchy and a set of geometric features and properties of the parts. The result of the transformation forms an X3D scene overlay over a camera view of the environment which can be seen in Figure 1. Note, that the superimposed text was generated by inclusion of the text-transformation.

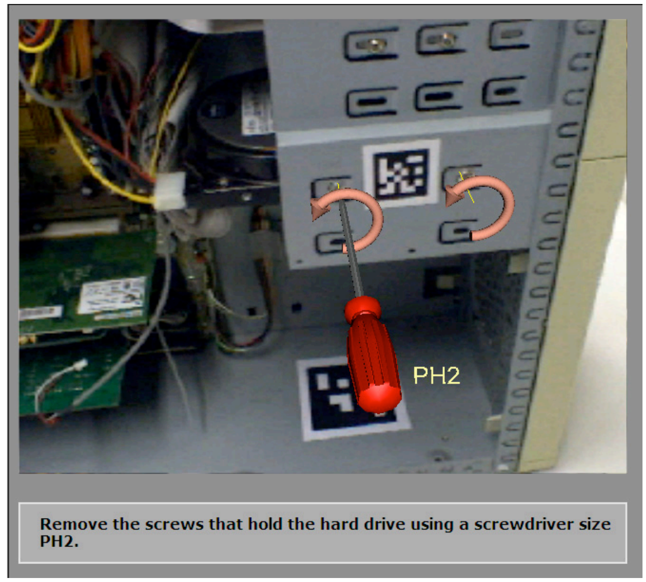

Figure 1. Screenshot of a graphical representation of the instruction shown in Listing 1.

Where the AR scene presented in Figure 1 is oriented more towards a hand-held augmented reality set-up, with the camera view embedded within a larger information display that provides supporting textual information and additional interaction, Figure 2 shows a slightly different representation of the instruction. Here, the supporting text is presented in form of a transparent overlay on top of the camera view, as it is required for use in head-worn augmented reality set-ups. Furthermore, the tool is displayed off-site in a separate tool area, which might improve clarity and perspective in certain scenarios.

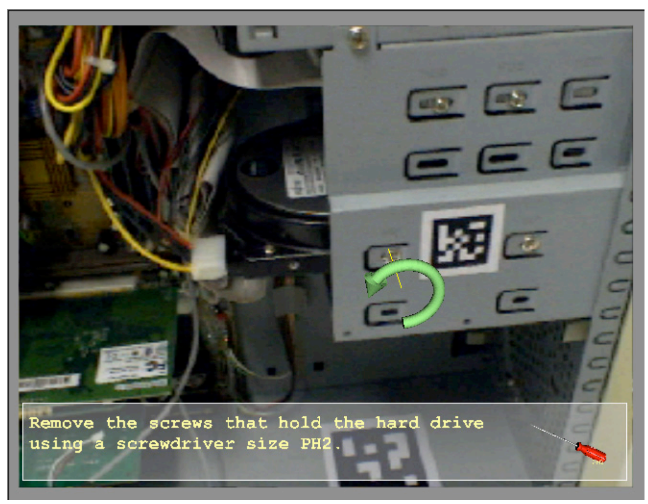

Figure 2. Small variations in the transform process offer the possibility to generate scenes in different visual styles.

The two scenes as well as the transformations they result from only differ in a few details. Their example shows, however, that minor variations of the transformation process offer the possibility to also generate documents in distinctive styles.

6 CONCLUSION

We presented our work on the generation of instructional AR-scenes from formalized technical instructions. Formalization was achieved by designing an abstraction model for instructive sentences based on a controlled natural language concept combined with spatial references to the involved entities. We specified an XML-based description format that has been designed to accommodate those concepts and explained the methods to generate visual and textual representations from those formalized technical instructions. Furthermore, we proved our concept with the help of a demonstrative example.

Further challenges for our research work will be to:

- Achieve an acceptable degree of completeness in the implementation of ASD-STE100 terms within the abstract instruction model. The presently implemented vocabulary does only cover a few test cases and examples.
- Implement an authoring tool that supports the author of the instructions by means of text- and menu-based predictive interface techniques.
- Increase the flexibility of visualization and to enable the definition of distinctive visual-interactive styles for the generated AR-scenes by implementing an additional stylesheet layer between transformation and visualization.

7 ACKNOWLEDEMENTS

The work presented is part of the AVILUSplus [2] project funded by the German Federal Ministry of Education and Research (ref. no. 01IM08002F).

REFERENCES

[1] ASD Simplified Technical English, Specification ASD-STE100, A Guide for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language, Issue 3, January 2005.

[2] AVILUSplus: Applied Virtual Technologies with Long-Term Focus on the Life Cycle of Products and Production Facilities. http://www.avilusplus.de.

[3] Cortona3D: Cortona3D RapidManual. http://www.cortona3d.com/interactive_electronic_technical_manuals.aspx

[4] Haringer, M., Regenbrecht, T. "A pragmatic approach to Augmented Reality Authoring", in Proceedings of ISMAR 2002, Darmstadt 2002 Pages:237 – 245

[5] Knoepfle, Ch., Weidenhausen, J., Chauvigné L., Stock, I.: Template based Authoring for AR based Service Scenarios. In Proceedings of IEEE VR 2005, Mar. 2005, Bonn/Germany, pp. 249-252.

[6] Kress-Gazit, H., Fainekos, G. E. and Pappas, G. J.: From structured English to robot motion. In IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, October 2007.

[7] Ledermann, F., and Schmalstieg, D.: APRIL: A High-level Framework for Creating Augmented Reality Presentations. In Proceedings of the IEEE Virtual Reality 2005 (VR'05), pages 187–194, 2005.

[8] MacIntyre, B., Gandy, M., Dow, S., and Bolter, J. D.: DART: a toolkit for rapid design exploration of augmented reality experiences. In Proceedings of the 17th Annual ACM Symposium on User interface Software and Technology (Santa Fe, NM, USA, October 24 - 27, 2004). UIST '04. ACM, New York, NY, 197-206.

[9] Martignoni, A. J. III and W. D. Smart: Programming robots using high-level task descriptions, in: Proceedings of the AAAI Workshop on Supervisory Control of Learning and Adaptive Systems, San Jose, CA, pp. 49–54 (2004).

[10] OMG: Semantics of Business Vocabulary and Business Rules (SBVR), v1.0 (2008)

[11] Schwitter, R., Tilbrook, M.: Controlled Natural Language meets the Semantic Web, in: A. Asudeh, C. Paris, S. Wan (eds.), in Proceedings of the Australasian Language Technology Workshop 2004, Macquarie University, 8th December 2004, pp. 55-62, 2004.

[12] Schwitter, R. and Tilbrook, M. 2006: Annotating websites with machine-processable information in controlled natural language. In Proceedings of the Second Australasian Workshop on Advances in ontologies - Volume 72 (Hobart, Australia, December 05 - 05, 2006). M. A. Orgun and T. Meyer, Eds. ACM International Conference Proceeding Series, vol. 238. Australian Computer Society, Darlinghurst, Australia, 75-84.

[13] Stock, I., Weber, M., Steinmeier, E.: Metadata based Authoring for Technical Documentation. In Proceedings of SIGDOC´05, Sept. 2005, Coventry/United Kingdom, pp. 60-67

[14] Wöhler, T., Wolter, L., Rothenburg, U., Strauchmann, M., Mader, S., Haupt, M.: Konzepte und Implikationen zur Integration von VR-, CAx- und PDM-Systemen für die funktionale Absicherung von Montage/Demontage-Operationen. In Proceedings of 13. IFF-Wissenschaftstage 15.-17. Juni 2010, Track 1 "Digitales Engineering und virtuelle Techniken", June 15-17 2010, Magdeburg/Germany, pp. 238-242.

[15] Zauner, J., Haller, M., Brandl, A., Hartmann, W.: Authoring of a Mixed Reality Assembly Instructor for Hierarchical Structures." In Proceedings of ISMAR 2003, Oct. 2003,Tokyo/Japan, pp. 237-246.