MULTIDOC – Multilingual Document Processing

LE3-4230 Final Report

IAI Saarbrücken   **September 1997**

Identification Sheet

Project ref. no.	*LE3-4230.*
Project title	**MULTIDOC – Multilingual Document Processing**
Deliverable status	*Public*
Contractual date of delivery	*September 1997*
Actual date of delivery	*September 22, 1997*
Deliverable number	*Final Report*
Deliverable title	*MULTIDOC Final Report*
Type	*REPORT*
Status & version	*Final 1.0*
Number of pages	*18*
WP contributing to the deliverable	*All*
WP / Task responsible	*IAI* *Martin-Luther-Straße 14* *D-66111 Saarbrücken*
Author(s)	*Rita Nübel and Jörg Schütz*
EC Project Officer	*Pierre-Paul Sondag*
Keywords	*automotive industry, multilingual documentation, language technology, product data, translation engineering, technical data management, integrated product development, intranet, extranets, object-centred design, CORBA, language control, controlled language, multilingual terminology, ontology, abstract documentation factory, software agents*
Abstract (for dissemination)	*This MULTIDOC report summarises the main findings and results of the MULTIDOC Concerted Action. The main result is the defintion of a so-called virtual application that can be seen as the compromise between the existing product documentation value cycles within the different MULTIDOC automotive companies and the MULTIDOC vision of the Abstract Documentation Factory facilitated through Translation Engineering. To allow for the development of both directions they are based on standardisable language resources, in particular they are based on a multilingual terminological ontology with multimedia capabilities* *In addition, this report gives an overview of the project timetable and discusses the internal evaluation and assessment of the implementation of the project.* *The conclusions of the report outline a project synthesis including farther business perspectives and a possible exploitation planning.*

Executive Summary

MULTIDOC is a European project based on the specific needs and requirements expressed by five representatives of the European automotive industry with focus on the multilingual aspects of product documentation. The general goal is to define and specify methods, tools and work-flows supporting stronger demands on quality, consistency and clarity in the technical information, and shorter lead times and reduced costs in the whole production value cycle of documentation including the translation into multiple languages. The results of the project will be applicable to any other component or system manufacturing business; thus, they are not restricted to the automotive industry.

The aim of the MULTIDOC Concerted Action was to identify the problem areas and to specify solutions for the European automotive industry when it comes to multilingual product documentation and also set a roadmap for the future. In software engineering, this phase is usually called the inception phase of an iterative software development process. During *inception* we establish the business rationale for the project and decide on the scope of the project. This is also the phase where we get the commitment from the project sponsor(s) to go further; in our case this was the successful evaluation of our MULTIDOC (Multilingual Documentation) project proposal which will start by the end of 1997 (LE4-8323).

The Concerted Action also included parts of the elaboration phase of a software development project. In *elaboration*, we collect more detailed requirements, do high-level analysis and design to establish a baseline architecture, and create the plan for *construction* which is the actual software production phase consisting of many iterations. In our domain, the most crucial bottlenecks concern the following business areas that need further elaboration:

- More and more languages in which product documentation has to be published; there is a drastically increased focus on Asian and East-European markets.
- Increasing costs for translations.
- Lead times in the translation process.
- Poor or no possibility to measure and control the translation process.
- Inconsistent use of information structure and information content.

All project partners agree that besides the quality of the product the services associated with the product and the accompanying documentation of the product must be seen as an integral part of the product. To satisfy the demand for high-quality technical documentation, the documentation has not only to be comprehensible and up to date, it has to be produced and delivered (including the accessibility to new or up-dated information) with modern technologies. The following scenario shall exemplify the direction:

Mr M is the proud owner of a new, environmentally clean car which was assembled according to his wish list from a huge variety of car components of the automotive manufacturer. The first contact with his new car was well before the actual delivery in a virtual reality animation, where Mr M was able to check his colour selection, the harmonisation of the chosen colour with the selected materials of the car interior, as well as first virtual driving tests.

Mr M is also very satisfied with the delivered car documentation; he got the personalised documentation right after the deal was contracted. This documentation is not only personalised but also customised to his specific car: this includes the appropriate colouring of all graphics in the documentation and the text itself, where we do not find any generalised references such as "... applies to specific countries.", "... according to the model variant." and so forth.

With the CD-ROM edition of the documentation Mr M can directly search for information on his PC at home, and he is also provided with multimedia enhanced information about his car. This documentation is also available in his car via the on board computer. This computer maintains each service work and possible repair work in its storage, and therefore permits customised service and repair measures and fault tracing at dealer's workshops. Not to mention, that these data are also available via the world-wide computer network of the car manufacturer.

Mr M gets new information about his car and about new products and services of the car manufacturer via his Internet access at home; for this he has subscribed to the free information service which in addition is parametrisable according to his information demands. Now, on a business trip to the south of Spain Mr M has to stop with a defect late at night. The workshop that is alarmed due to the 24-hours assistance service is somehow lost: this kind of defect is not listed in the technical service documentation and therefore not a standard service and repair routine. However, the immediate access to the hotline information service via the computer network identifies the same defect two days ago in Oregon/USA. The problem solution that the Spanish mechanic gets on his workshop screen is in English. So he activates the translation-on-demand button and receives a Spanish translation within a few seconds. This translation is not perfect, but the necessary repair steps and the terminology of the needed tools and parts is correct due to a multilingual terminology knowledge base which is maintained by the car manufacturer; therefore the mechanic does not care about the 'ser/estar' errors of the delivered Spanish text.

Not every service information is available in all languages that are supported by the car manufacturer; the translation of service information is a matter of information need, but every available information is accessible through "pull" technology (for example, the hotline information service). Information of common interest is distributed in all language through "push" technology; this ensures a fast and efficient update of all product documentation. The defect of Mr M's car is now available in English and Spanish (after a correction of the computer translation); after an in-depth analysis of the defect this information might become available in all supported languages.

This scenario, which of course is partly a vision, shows the necessity of integrating services, documentation and networked IT solutions with the support of modern, multilingual language technology.

It is the basis of the virtual application of MULTIDOC that constitutes a compromise between the present situation of product documentation in the different automotive companies and the MULTIDOC *vision* of an *Abstract Documentation Factory* based on the concept of Translation Engineering. This virtual application is the result of the elaboration phase and allows for a smooth and cost effective transition of the business. Thus, we have first and foremost concentrated on the existing process stages of technical documentation, where several control capabilities for the source language may support the technical writer and other knowledge workers in identifying and defining information objects in an SGML authoring environment. In broad terms, an *information object* is either a meaningful, non decomposable SGML marked-up text unit or a composition of such text units. The virtual application shall already include steps toward Translation Engineering that is the operational foundation of the MULTIDOC vision. *Translation Engineering* (TE) as a business strategy is concerned with:

- Fostering the use of information objects preferably linked with product data to ensure the consistent use of information structure and information content.
- Optimising the translation production chain through the employment of different multilingual language technologies, including multilingual generation.
- Linking of source language information and target language information to facilitate better maintenance, quality assurance and quality control.

Translation Engineering as a methodology for multilingual documentation will help to drastically reduce cost, to shorten lead times and to further improve the quality of technical documentation. We talk about a *virtual application* because this application scenario is based on the generalisation of the different documentation work-flows maintained at the automotive user sites of the MULTIDOC partners.

In addition, the core business and the component or system manufacturing industry benefits in terms of:

- accelerating the building of enterprise-wide and industry sector wide knowledge systems (repositories and knowledge bases) based on Web technology (intranets and extranets) including multimedia (text, graphics, video and animation, virtual reality) and multimodal (language and speech navigation) capabilities,
- improving the semantic content of information objects (accuracy and quality),
- speeding up of the translation processes (today, in limited cases the translation process could be substituted by multilingual generation and *symbolic authoring*, which in particular has to be seen in combination with *controlled languages*),
- reorganising of the overall production process (*lean multilingual documentation*).

In the Virtual Application report ([MD-3, 1997]), we have introduced the MULTIDOC vision of the Abstract Documentation Factory (ADF) based on the concept of TE. TE as such will revolutionise the current way of thinking in technical documentation because the whole documentation process is oriented toward multilingualism. This new business scenario includes a push/pull policy for technical information delivery and retrieval in an automotive dealer's workshop in combination with a translation-on-demand policy (see the scenario above). TE is responsive to the new business demands which are concerned, on the one hand, with an increase of languages for product documentation, in particular Asian languages and East-European languages, and on the other hand, with the reduction of documentation cost, lead times and efficient measures to control and manage the translation tasks. TE as a methodology for multilingual documentation will harmonise and unify the most crucial documentation requirements in areas such as the consistent use of technical information in structure and content, the efficient and effective reuse of information objects based on standardised information structures (increased retrieval hit-rates), and the terminological and multilingual orientation of the whole information production process.

Since the MULTIDOC vision is a more long-term development tasks we analysed the existing documentation processing chain and we identified the stages for initial improvements taking into account the historical evolution of product documentation. This led to the definition of a strategy for an efficient and effective employment of language technology to bridge the gap between the present situation in product documentation and the MULTIDOC vision. This strategy is based on the present situation and has to be maintained with various restrictions for the different automotive companies but with the common interest to work toward the ADF vision that is shared by all companies, however, with different ways to reach the vision.

Whereas the realisation of the "bridging the gap" strategy is scheduled for a two to two and a half years development plan, the vision will take approximately five years These are additional two and a half years development, because both directions are based on a common foundation: the MULTIDOC *ontology* that constitutes an advanced terminological knowledge base with multimedia capabilities.

All development strategies have been validated with a cost/benefit appraisal based on a hypothetical business calculation of a virtual automotive enterprise. We have taken this way to further maintain the generalisation direction which we already followed in the other phases of the Concerted Action. However, our profitability assessment is based on actual calculations made by the MULTIDOC partners for their specific enterprise situation. The validation stage also included a risk analysis consisting of:

- Critical analysis of our approach.
- Analysis of changes to the human resources, the technical infrastructure and the organisational environment.
- Identification of any restrictions, constraints, risks and problems hitherto not taken into account.

The user groups who are in the focus of our work are, on the one hand, the knowledge producers of documentation departments and translation departments (technical writers and translators, designers and engineers, and so forth) and the knowledge consumers in the automotive workshops (mechanics and technicians).

Table of Contents

1   Project Timetable

1.1   Contractual

Due to a professional project management the implementation of the project was strictly guided by the workpackage definitions specified in the Technical Annex. All tasks were successfully fulfilled within the foreseen timeframes and the allocated budget.

However, it should be mentioned that sometimes the communication and information flow between the partners could have been more efficient, but often this was a matter of time constraints of the industrial partners.

1.2   Stages of Work

The work carried out was divided into three main stages:

1. Analysis of the present situation in product documentation, including the identification of existing bottlenecks as well as the demands and requirements of different user groups (knowledge producers and knowledge consumers).
2. Market analysis to identify already existing solutions and commercial of the shelf (COTS) products.
3. Specification of a MULTIDOC solution that included a vision and roadmap strategy for its realisation and a cost/benefit analysis of this solution.

Although the analysis of the present situation was defined as a self-contained workpackage it was dealt with as a continuous task, and thus, it contributed to a common understanding of what is feasible today in the field of IT and LT and the directions of tomorrow. This then is reflected in the MULTIDOC vision which presents an agreed future direction of multilingual product documentation.

The market analysis gave the partners new insights into the expanding documentation market which first and foremost is concerned with enterprise-wide documentation management systems that all lack the integration of language related functionality.

The MULTIDOC solution is an agreed direction for the follow-up project. The envisaged work will distinguish between general developments such as the terminological ontology and partner-specific developments which are concerned with on-site integration or interfacing of the implemented language technology functionality such as the communication with existing document management systems or SGML editors.

1.3   External Reviews

Since we did not had an official external review, we organised our own reviews with the partners and in particular with the new partners that joined the consortium. This strategy also opened new or even different views on the MULTIDOC solution.

1.4   Conferences, Exhibitions and User Group Meetings

The project was presented at several international conferences, concertation meetings of the European Commission and User Group workshops and technical meetings with external participants from the automotive industry, the manufacturing industry, and the IT and LT industry (see also the list in the Annex).

1.5   Other important Events

The project maintains a World Wide Web home page at URL **http://www.iai.uni-sb.de/multidoc** which contains general information about the project in three languages (English, French and German), the executive summaries of all project deliverables, a list of MULTIDOC related events and services including a registration form for the MULTIDOC

Interest Group which is an integral part of the MULTIDOC Network.

2 Achievements

2.1 Demonstrator System

Although the project was not obliged to deliver any software system, it was felt that the work would be improved and the understanding would be accelerated by software presentations. In collaboration with the German MULTILINT project, we developed additional LT modules for English and French (mainly in the terminology field and the style checking area) which can be demonstrated under the existing MULTILINT user interface.

In addition, a first prototype of a multimedia terminology component was realised. This component allows the linking of terms to graphics and pictures in both directions with a hierarchical navigation functionality. This means, the user may select a term during the authoring process in his editor and the associated graphical representation is displayed. Since all graphics are linked with the same partial order as the domain terminology the user can navigate either within the graphical representation or within the terminological representation.

This implementation work also gave us new insights into the integration and interfacing problematic which often is underestimated by system developers, and it will form the implementation basis of the follow-up project.

2.2 Research Results

Major achievements of this one-year Concerted Action have been reached on different levels or strategic areas, where one level closely interrelates with the other. The level which has initiated and guided all subsequent research activities in the CA deals with the investigation and analysis of genuine user needs and requirements in the field of technical documentation. It is therefore viewed as the backbone or basic level on which the activities have been concentrated in the beginning of the project and which pervades the successive levels of investigations. A more general view on the complex situation of technical documentation and related problems in several technical domains has been taken on the business-oriented level. Here, the project activities concentrated on the description of the present market situation, the identification of business needs in this sector and the current and future trends. Taking into account the user needs analyses and the business situation identified on the preceding levels, a scenario has been designed on a conceptual and future-oriented level which proposes a possible solution to the present problems of technical documentation.

2.2.1 Requirements Analysis

On the basic level, the MULTIDOC user requirements have been identified within different domains. From the beginning, the user requirements analyses have been considered as the MULTIDOC strategic essentials and as the backbone of further investigations. This view has been supported by the ultimately user-guided approach of MULTIDOC (UCD, User-Centred Design).

Initial user needs and requirements have been identified and analysed within several strategic domains including:

- User orientation in terms of profiles, tasks and work-flows.
- Standardisation and normalisation specific to the industrial sector such as SAE J1930 (terminology), SAE J2008 (DTDs), CALS and STEP (product data), as well as those standards related to product documentation and its IT environment such as SGML and HyTime, UML, CORBA, Java.
- Technical infrastructures including advanced IT in the areas networking, client/server, database and object-oriented technology.

- Linguistic intelligence, including, multilingualism, in areas such as communication, translation, and the general LT infrastructure.

These needs and requirements are initial insofar as the identification and analysis comprise iterative cycles between automotive product data environments, product documentation environments and LT/IT service and tool providers and suppliers.

Two types of users have been identified throughout the user analysis which comprised the set-up of user profiles and related tasks. Language Technology makes multilingual documentation a more effective information medium on the one hand for *knowledge workers* (technical writers and translators), who are responsible for the production of documentation, and, on the other hand, for *knowledge consumers*, who range from workshop service and repair staff, marketing and sales personnel, to product owners.

As a first step toward standardisation, one particular application domain was selected: *service and repair methods.* We have taken this domain to elaborate the whole production chain of multilingual product documentation. This selection is mainly inspired by the kind of product data that technical writers receive from design and construction departments in the form of diagrams or skeletons. Additionally, it allows to pursue the whole production chain from product data to publishing, and the same information can be re-used for different purposes, for example, the training material for workshop mechanics.

Within actual work-flows, multilinguality has an outstanding position. Most enterprises have outsourced the translation of their documentation since they do not have the infrastructure nor the personal resources and competence required for this complex task. However, the delegation of translation activities to translation agencies outside an enterprise means a loss of control in terms of quality as well as cost/benefit relations. Additionally, the synchronisation of source language documents and their translations becomes an organisational challenge. Therefore, early-integratable LT support is required for the source documentation chain which extends to the control of the translation process, e.g. by integrating a multilingual terminology knowledge base which is also used by the translation departments. A prerequisite for this approach is of course a flexible IT architecture which allows for an effective control of the complex process of multilingual documentation in terms of terminological, grammatical and style consistency and correctness, the support of the translation process, and the facilitating of full text information retrieval capabilities.

Similar requirements hold for the publishing process, where an integrated approach is required for the diverse tasks of formatting and printing information according to the envisaged output medium such as DTP, print, on-line documentation, CD, microfiche, archive, and electronic exchange.

2.2.2 Market Research

On a business-oriented level, an analysis of the present complex situation in the field of technical documentation in the component manufacturing industry has been carried out. A concise description of the problem domain and the state-of-the-art has been provided, and the future trends to be observed in this area have been identified.

The management process of any technical information during its whole life-cycle has to fulfil the business requirements *speed*, *efficiency* and *quality*. This means that the quality of the information associated with a product or with a component/part of the product in its various manufacturing and usage stages and independent of its use has always to be measured along these lines. The role of technical information is no longer restricted to just formally fulfil the purpose of documentation only. Rather, more and more enterprises come to understand that technical information is emerging as the dominant pivot for organising and presenting corporate knowledge.

2.2.3 Vision and Roadmap

Building on the results achieved in the previously described levels, further investigations of the MULTIDOC CA have moved to a conceptual level, where the project work has brought about the development of a future-oriented scenario for multilingual technical documentation: The Abstract Documentation Factory (ADF). The central ideas of ADF are described in the Virtual Application (see next section).

As a first move towards the realisation of the ADF Vision, we have designed the MULTIDOC Roadmap. It provides a development strategy in order to bridge the gap between the visionary aspects of ADF and the present situation in technical documentation environments which calls for integratable short-term solutions. It serves as a link to the future-oriented concepts of ADF by anticipating these new ways of document production and by providing short term solutions which can be integrated with existing business procedures. This approach permits an early validation through both the users and the enterprise's management.

The guiding principle of the Roadmap approach is that LT components are integrated as early as possible in the existing documentation process. Early integration of linguistic-based control functionality helps improving the content quality of source language *information objects* (self-contained units of information which may be composed of smaller, atomic, i.e. non-decomposable information units) on different (linguistic) levels and in the various stages of documentation production.

Improving the linguistic quality of source language information objects means also improving the quality of their translation into different languages. In this scenario, the aspect of linguistic control of source language documents stands in the foreground. Nevertheless, there is already a strong orientation towards the translation process implicit. At this stage, LT with a view to translation engineering (TE) is still a support utility, thus supporting the production of documents without affecting too much aspects of reorganisation and restructuring of processing steps (or the processing chain as a whole).

The diverse linguistic levels of control build on a multilingual terminological ontology, which is considered as the backbone of any kind of technical information to be managed. It is thus necessary that the product data is readily linked with such an instance of a term bank (mono- and multilingual), and that consistency and coherence checks can be performed with the term bank's data. This term bank should be part of a central, multimedial knowledge base (termi-nological repository) with a broker component that permits search, retrieval and navigation functionality.

Linguistic-based control functions relate to terminology consistency, spell and grammar checking as well as the application of writing rules. Writing rules specify style conventions of how to distribute and structure information on a phrase and sentence base (style checking).

Finally, in this scenario, where the first and foremost goal is to increase the quality of documents by introducing LT at all possible stages of the processing chain, the term "translatabil-ity" comes to play a crucial role. All LT functionalities as already described contribute to the translatability of documents, either for human translation or for different types of machine translation (high-quality or raw machine translation) as well as for Translation Memory tech-nology. Thus, the Roadmap scenario already provides implicit mechanisms for a better coor-dination of source language documents and their translation by linking them indirectly through the application of linguistic-based control functions.

2.2.4 Virtual Application

We introduced the MULTIDOC vision which is concerned with Translation Engineering (TE) as a means for reorganising the workflow and value cycles of multilingual product documen-tation. We elaborated the basics of TE and established the baseline architecture of the Abstract Documentation Factory (ADF). This work is based on the collection and the analysis of de-tailed requirements and the identification of the risk categories that concern:

- Requirements risks that concern the verification of our understanding of the business demands.
- Technological risks that concern the evaluation of. the feasibility of our technology foundation.
- Skills risks that concern the selection of the right personnel for the realisation.
- Political risks that concern the conformity of our solution with the existing corporate identities of the industrial partners, which, however, is a matter to be solved by each of the involved partners.

The initial requirements analyses as reported in [MD-D1, 1997] where further developed in form of use cases which are snapshots of certain aspects of the intended system solution and a basis for the communication with the domain experts. The baseline architecture of the ADF then was founded on:

- List of use cases.
- Domain model, which captures our understanding of the business and serves as a starting point for our key domain concepts (classes).
- Technology platform, which describes the key ingredients of our implementation technology and how they fit together.

This architecture is the guide of our construction phase; it acts as the blueprint for all later stages. Inevitably, the details of the architecture will change, but it should not sustain too many serious changes. It is envisaged to further elaborate the ADF approach during the construction phase of the "bridging the gap" modules and components, as described in [MD-D2/D4, 1997] and farther elaborated in [MD-D3, 1997]. Both approaches are centred around the MULTIDOC terminological ontology as the primary information source. The parallel development allows for an optimal use of resources, and permits a straight forward implementation of the ADF based on already existing LT modules and components. This is also the reason for our time-to-market estimate of additional two and a half years after the completion of the "bridging the gap" toolbox. It should be noted that the ontology centric approach has several benefits which have an direct impact on the most important business demands in multilingual documentation. Among them the most significant ones are that it

- supports harmonisation and standardisation between product data environments and product documentation environments;
- ensures better control of information object production, translation and deployment because of clearly defined responsibilities and quality assurance measures, including a transparent work-flow control;
- fosters a better integration of supplier information and subsidiary information (core business as well as associated businesses);
- abandons the need of an end-control within the information object production value cycle because of the distributed responsibilities with an integrated control functionality;
- fits with existing and emerging networked computing environments, including advanced agent technology.

2.3 Other Results

Besides the results briefly described in the previous sections, the project benefited very much from the collaborative efforts of the many skills and experiences which for the first time came together in an industry oriented European project. Both industry and research partners learned from each other with the result of an enabling system's solution for multilingual product documentation. As an example, the work on "controlling language vs. controlled language" and "automatic, multilingual natural language generation" shall be mentioned.

3 Evaluation and Assessment

3.1 Verification and Validation

The project partners have evaluated the cost/benefit ratio we gain through the introduction of language technology into the documentation processes and the coupling of LT with advanced information technology. Two return on investment scenarios for the calculation of a cost/benefit ratio were developed. In the first scenario, we elaborated a profitability assessment for the introduction of LT in the documentation work-flow. The second scenario, evaluated the cost/benefit balance on the basis of a more effective documentation terminology management taking into account multilinguality as early as possible.

Since the introduction of LT tools into the documentation work-flow can be done in a step by step manner, the technology can first be verified and tested in selected domains and with well trained personnel. This, however, shall be done as soon as first software releases are available to open the possibility to take care of requirements which have not been communicated during the elaboration phase. In so far, LT as such does not have any specific risk potential, although the way authoring could be defined in future work-flows, especially when adopting a strict information object centred view, may have a dramatic influence on psychological factors of the technical writers.

A real risk potential is the integration or the interfacing of the LT tools with the existing IT infrastructure. This in particular, because the companies are in a re-organisation phase with the introduction of full native SGML environments. Today, there are still in between solutions, for example, the use of non-native SGML documentation tools with a structure-oriented WYSIWYG editing mode. To avoid a wrong integration path, we should thoroughly elaborate alternative strategies for the integration. This could be a Web browser based interface for the presentation of the results of the LT tools. In this case, the technical writer would be responsible for the appropriate correction in the actual document, which would not be under the direct control of the LT tools. Such a solution, however, would also prepare the road to the MULTIDOC vision of the ADF.

The terminology orientation, especially with a multilingual view, will have a significant influence on all stages of the documentation work-flow. On the one hand, it is the foundation of a better quality documentation because it ensures consistency on the technical and on the language level (content), and on the other hand, it is a means for reducing the costs both in monolingual documentation and in multilingual documentation (translation tasks fulfilled by suppliers). The changes based on terminology will also open the way to viewing documentation as an inherent part of the automotive business, and thus, contribute to future changes in the direction of the ADF.

The alternative strategy of entirely outsourcing the technical documentation should be considered seriously with all its pros and cons. This, however, must be based on a thoroughly established communication platform between the automotive core business and the multilingual documentation supplier.

Further details in this direction, however, were beyond the scope of this project. Nevertheless, the elaboration of a business strategy for this approach could be the subject of a separate feasibility study within the automotive industry. The strategy should be based on similar mechanisms and legal foundations as it is the case with today's automotive suppliers within the core business.

3.2 Internal Reviews and Feedback

The business scenarios in product documentation as they exist at present are the result of an evolutionary process into which additional processes or organisations, such as the SGML environment, online publishing in different forms and formats, Asian languages support based

on Unicode and so forth, were/are integrated with new demands such as the SAE J2008 standard, technical information delivery on compact disks and the World Wide Web, new markets for the product, and so forth.

To keep the cost for such new services affordable, many of these new processes were outsourced to competence centres or in-house profit centres were established. This gives the enterprises the freedom to concentrate on the core business and not to mix competences. Documentation related activities could thus be managed similar to the services provided by automotive suppliers where strong relationships are defined according to the lean production methodology. This, however, is not the case in the present documentation situation, and therefore has led to an increase of translation cost and translation time. In addition, the situation is farther complicated by the fact that there are no bilateral established measure and control procedures for the whole translation value cycle.

Often, the translation subsidiaries and suppliers are responsible for checking the source language documentation for consistency and translatability (pre-editing process according to the corporate standards of the customer) and the building and maintenance of multilingual glossaries including corporate terminologies to support the translation process besides there core business that is concerned with the documentation translation proper. Today, this core business concerns the management of the translation process (including the organisation of a translators network for languages that are maintained), the building and maintenance of translation memories and the post-editing of the translated documentation, including quality assurance measures at each processing stage. In addition, translation subsidiaries have to accept a variety of input and output formats, and they have to maintain heterogeneous platforms according the equipment of their customers (multi-platform, multi-format documentation). With the introduction of SGML-based documentation this situation has improved. Today, the common criteria for all translation services are:

- lead time optimisation,
- cost optimisation,
- application of advanced technologies including leading edge technology,
- flexible capacity management,
- quality assurance measures at all stages of the translation process.

Here, a better co-ordination between the automotive customer and the translation supplier would also help to improve the present situation. This could be a better quality of the source language documentation and a joint effort in the field of multilingual glossaries and terminologies, as well as agreed standards when it comes to quality assurance and the reuse of existing multilingual information objects.

There are possible alternative strategies than the ones we have identified within the MULTIDOC elaboration phase, where we have focussed on the documentation value cycle as an intrinsic part of the automotive core business, i.e. documentation as yet another component of the manufactured product.

Since the outsourcing of certain services and the production of car parts is a common practise in the manufacturing business (automotive suppliers build a whole business branch), we also could envisage a documentation strategy that is based on this outsourcing principle, i.e. the production of documentation is entirely done by a documentation supplier (this should be a multilingual documentation supplier). The organisation of such a service should be similar to the automotive supplier services, meaning that effective co-ordination and efficient exchange of information is the basis for a successful fulfilment of the documentation tasks.

An ontology which harmonises and standardises the product data environment and the documentation environment can be seen as the pivot for such a business strategy. On the one hand, it would serve as the corporate backbone of a company's knowledge resource, and on the other hand, it would be the authoritative information source for any kind of technical documenta-

tion. The multilingual documentation supplier would then be responsible for the building and maintenance of the ADF. One obvious benefit is the better separation of responsibilities which also avoids the mixing of competence domains. However, this approach needs further elaboration since we also have to take into account the competitive advantage that a company gains through good documentation when it comes to service and maintenance operations of their products.

3.3 Internal Collaboration

The project collaboration was established at three levels which were all supervised by IAI:

1. Internal working groups
2. External consultancy
3. Workshops

The internal working groups worked on selected fields (agreed upon in technical meetings) and elaborated possible solutions which were communicated to the other project participants. The results of this group were normally discussed at technical meetings and documented in the deliverables of the project and in internal documents such as meeting minutes.

External consultancy (mainly in the field of IT but also LT) was used within all project stages; this since most IT and some of the documentation work, especially translation tasks, are outsourced to specialised companies.

Workshops were used to communicate the project results to a wider audience and to discuss alternative system solutions and strategies. Each workshop had a specific focus: general user requirements, terminology, and language control.

4 Conclusions and Future Prospects

The long-term goal of MULTIDOC is the definition of a TE methodology and a TE strategy which gives up the present way of thinking that the documentation processing chain is strictly chronological or linear, not linked with product data environments, and translation being a separate process at the end of the processing chain. As a first step in this direction industrial R&D is working on an approach that integrates the STEP world with the SGML world as a basis for building the needed information system's infrastructure. Our approach, however, concerns in addition the combination of intelligent LT with sophisticated IT to build this information system. This approach still would support a linear view on the documentation value cycles in that LT is integrated step-wise in the existing documentation work-flow. Since a more translation-oriented or multilingual view is aimed at, a solid basis from which the vision of the ADF may be further developed is already established. The ADF comprises the complete re-organisation of the documentation processes: vertically and horizontally. This means that LT in general, and specifically multilingual LT including translation technology, will make the move from a supporting technology to an enabling technology. The most important areas for this development are:

- Graphics and other multimedia incarnations, such as video, animation and virtual reality applications, may enrich or even replace text in certain information objects and facilitate new approaches to information production such as symbolic authoring.
- Translation-on-demand policy to allow for an efficient and effective control of the actual translation needs, because not all information objects need to be stored in every language that is supported by the business (translation management).
- Compilation of documents from multilingual information objects, either already stored in a foreign language, translated on demand, or generated from an abstract

representation; this allows for the simultaneous delivery (publishing) of multilingual documentation.

LT makes multilingual documentation a more effective information medium for knowledge workers (technical writers and translators), who are responsible for the production of documentation, and for knowledge consumers, who range from workshop service and repair staff, marketing and sales personnel, to product owners. It may also change the existing working habits and working routines of the different types of users. Combined with sophisticated information technology multilingual documentation gets even more effective in terms of information dissemination and information retrieval. Our focus will be on the following three main components:

1. Ontology as a means for representing domain knowledge (the subject of technical documentation) linked with natural language semantics.
2. Object Modelling Technique (OMT) as a theoretical foundation for analysis and design, and as an implementation platform based on, for example, CORBA.
3. Agent technology as the overall umbrella for construction, and as an alternative implementation platform, especially for networked applications.

In the ADF, knowledge producer and knowledge consumer will operate in virtual environments brokered by software agents. A software agent acts autonomously on behalf of a person to fulfil the person's goal or task. Agents are also key enabler for push technology which is used in information update tasks and information retrieval tasks. Push technology is based on the idea of content customisation and tailored interaction (see also the scenario described in the Executive Summary).

5   References

[Bateman, 1992] John Bateman, 1992. The Theoretical Status of Ontologies in Natural Language Processing. In Proceedings of the International Workshop on Text Representation and Domain Modelling, TU Berlin, Germany.

[Bevan, 1996] Nigel Bevan, 1996. User-centred Design. Deliverable D3.1.2 v1.2 of IE 2016 project INUSE.

[Bradshaw (ed.), 1997] Jeffrey M. Bradshaw, 1997. Software Agents. AAAI Press/MIT Press, Cambridge, Massachusetts, USA.

[Card et al., 1996] Stuart K. Card, George G. Robertson and William York, 1996. The Web-Book and the Web Forager: An Information Workspace for the World-Wide-Web. In Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems, Vancouver, Canada.

[Church & Hovy, 1993] Kenneth Church and Eduard Hovy, 1993. Good Applications for Crummy Machine Translation. In: Machine Translation 8, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[Cockburn, 1997a] Alistair Cockburn, 1997. Structuring Use Cases with Goals. Web document at URL http://members.aol.com/cockburn/papers/usecases.htm.

[Cockburn, 1997b] Alistair Cockburn, 1997. Using "V-W" Staging to Clarify Spiral Development. Web document at URL http://members.aol.com/cockburn/papers/ vwstage.htm

[Drucker, 1946] Peter Drucker, 1946. The Concept of Cooperation. John Day, New York,

USA.

[Englander, 1997] Robert Englander, 1997. Developing Java Beans. O'Reilly & Associates, Inc., Sepastopol, CA, USA.

[Flanagan, 1996] David Flanagan, 1996. Java in a Nutshell. O'Reilly & Associates, Inc., Sepastopol, CA, USA.

[Flanagan, 1997] David Flanagan, 1997. Java in a Nutshell. Second Edition covering Java 1.1, O'Reilly & Associates, Inc., Sepastopol, CA, USA.

[Foyler, 1997] Martin Foyler, 1997. UML Distilled. Addision Wesley Publishing Company, Reading, MA, USA.

[Haller, 1996] Johann Haller, 1996. *MULTILINT - A Technical Documentation System with Multilingual Intelligence*. In Proceedings of ASLIB Translating and the Computer 18, London, England.

[Haller & Schütz, 1997] Johann Haller and Jörg Schütz, 1997. Integration linguistischer Intelligenz in die multilinguale technische Dokumentation. To appear in Proceedings of EUROMAP Forum „Sprache ohne Grenzen", Munich, Germany.

[Herdman, 1997] David Herdman, 1997. *No mirrors and magic here - You can write common code for Unix and NT*. Sunworld On-line, April issue and May issue.

[Howells, 1997] Ian Howells, 1997. *Document Management Beyond 2000: More of the Same, a Strategic Tool or Spider in the Web?*. In: Document World, Issue 1, IMC and AIIM publication, Powerhouse Solutions Ltd., Surrey, England.

[Huntington, 1996] Samuel P. Huntington, 1996. The Clash of Civilizations. Simon and Schuster, New York, USA.

[Maguire, 1997] Martin Maguire, 1997. RESPECT User Requirements Framework Handbook. Deliverable D5.1 of Information Engineering (IE) 2010 project RESPECT. HUSAT Research Institute, Loughborough, UK.

[MD-D1, 1997] Rita Nübel and Jörg Schütz, 1997. MULTIDOC-CA Requirements Reports. Deliverable D1 (LE3-4230), IAI, Saarbrücken, Germany.

[MD-D2/D4, 1997] Rita Nübel and Jörg Schütz, 1997. MULTIDOC White Paper: Trend Research, Vision and Roadmap. Deliverables D2 and D4 (LE3-4230), IAI, Saarbrücken, Germany.

[MD-D3, 1997] Rita Nübel and Jörg Schütz, 1997. MULTIDOC Virtual Application: The Abstract Documentation Factory. Deliverables D3 (LE3-4230), IAI, Saarbrpcken, Germany.

[Nwana, 1996] Hyacinth S. Nwana, 1996. Software Agents: An Overview. Intelligent Systems Research AA&T, BT Laboratories, Ipswich, United Kingdom. Also available on the Web at URL http://www.cs.umbc.edu/agents/introduction/ao.

[OMG, 1995] Object Management Group, 1995. The Common Object Request Broker: Architecture and Specification Version 2.0

[Orfali et al, 1996] Robert Orfali, Dan Harkey and Jeri Edwards, 1996. The Essential Distributed Objects Survival Guide. John Wiley & Sons, Inc., New York, NY, USA.

[Orfali et al, 1997] Robert Orfali and Dan Harkey, 1997. Client/Server Programming with Java and CORBA. John Wiley & Sons, Inc., New York, NY, USA.

[Ovum, 1996] Heather Stark, Ashim Pal and Rory Staunton, 1996. Ovum Evaluates: Docu-

ment Management. Ovum Ltd, London, England.

[Reese, 1997] George Reese, 1997. Database Programming with JDBC and Java. O'Reilly & Associates, Inc., Sepastopol, CA, USA.

[Roan et al., 1995] Ramana Roan, Jan. O. Pederson, Marti A. Hearst and Jock D. Mackingly, 1995. Rich Interaction in the Digital Library. In Communications of the ACM, Vol. 38, No. 4, pages 29-39.

[Schütz, 1994] Jörg Schütz, 1994. Terminological Knowledge in Multilingual Language Processing. Studies in Machine Translation and Natural Language Processing, Volume 5, CEC Luxembourg and Brussels, Belgium.

[Schütz, 1996a] Jörg Schütz, 1996. Combining Language Technology and Web Technology to Streamline an Automotive Hotline Support Service. In Proceedings of AMTA-96, Montreal, Canada.

[Schütz, 1996b] Jörg Schütz, 1996. Network-based Machine Translation Service. In: Proceedings of the EAMT Machine Translation Workshop, Vienna, Austria.

[Schütz, 1997] Jörg Schütz, 1997. Utilizing Evaluation in Networked Machine Translation. In Proceedings of TMI 97, Santa Fe, New Mexico, USA.

[Spence, 1995] Malcolm Spence, 1995. An Assessment of STEP as an Architecture for Product Data Integration. Digital Equipment Corporation.

[Sowa, 1984] John F. Sowa, 1984. Conceptual Structures: Information Processing in Mind and Machine. Addision Wesley Publishing Company, Reading, MA, USA.

[Sowa, to appear] John F. Sowa. Knowledge Representation: Logical, Philosophical and Computational Foundation. To be published by PWS Publishing Company.

[Sörman, 1997] Mats Hultemark, 1997. Personal communication. Sörman Information AB, Växjö, Sweden.

[Toyota, 1988] Toyota, 1988. A History of the first 50 years. Toyota Motor Corporation, Toyota City, Japan.

[Tucker, 1996] Hugh A. Tucker, 1996. Harmonization of SGML and STEP. ISO (TC 184/SC4/WG3/T14 Product Documentation) STEP/SGML Tutorial.

[UML, 1997] Unified Modelling Language (UML) Documentation, Version 1.0. Rational Software Corporation, Santa Clara, California, USA. Available on the Web at URL http://www.rational.com.

[Unicode, 1996] The Unicode Standard: World-wide Character Encoding. Addison- Wesley Publishing Company, Reading, MA, USA.

[Womack et al., 1990] James P. Womack, Daniel T. Jones and Daniel Ross, 1990. The Machine that changed the World. Rawson Associates, New York, USA.