

# Requirements Gathering

Requirements collection refers to the process of gathering and documenting the specific needs, goals, and constraints of stakeholders for a project or product. This information is used to guide the design, development, and testing of the final solution. Effective requirements collection involves collaboration with stakeholders, clear communication, and a thorough understanding of the problem being solved.

- **Here are the steps for effective requirements collection:**

1. Identify stakeholders: Determine who will be impacted by the project and who needs to provide input.
2. Interviews: Individual or group meetings with stakeholders to gather information and understand their needs.
3. Surveys: A written questionnaire distributed to stakeholders to gather information and feedback.
4. Workshops: Collaborative sessions with stakeholders to identify, prioritize and validate requirements.
5. Observation: Observing and recording the activities of stakeholders to gather information about their work processes and requirements.
6. Prototyping: Creating a functional model of the solution to gather feedback and requirements from stakeholders.
7. User stories: A narrative technique for describing the desired functionality and user experience of the solution.
8. Review of existing documents: Examining existing documentation such as business processes, user manuals, and competitor analysis to gather requirements.
9. Focus groups: Bringing together a small, diverse group of stakeholders to discuss and provide input on requirements.
10. Requirements traceability: Tracing the relationship between requirements and project deliverables to ensure they are met.
11. Stakeholder analysis: Analyzing the interests, influences, and expectations of stakeholders to gather requirements.
12. Business process modeling: A visual representation of business processes to gather requirements and improve efficiency.
13. Reverse engineering: Examining a current product or solution to identify requirements for a new solution.
14. JAD (Joint Application Development) sessions: Collaborative sessions with stakeholders and technical experts to gather requirements.
15. Mind mapping: A visual tool for organizing and categorizing information and requirements.
16. Root cause analysis: A problem-solving technique that helps identify the underlying causes of a problem to gather requirements.
17. Ethnographic research: A research method that involves observing and interacting with users in their natural environment to gather requirements.

18. Concept modeling: A visual representation of the key concepts and relationships involved in a project to gather requirements.
19. Performance testing: Testing the solution's performance under different conditions to gather requirements for performance optimization.
20. Competitive analysis: Examining the offerings and strengths of competitors to gather requirements for a unique solution.
21. Gap analysis: Comparing the current state of a system with its desired future state to identify and gather requirements for improvement.
22. Requirement prioritization: Prioritizing requirements based on their importance and impact to the project.
23. Cost-benefit analysis: A technique that evaluates the costs and benefits of requirements to make informed decisions on which requirements to pursue.

### **Abstract**

Numerous studies in recent months have proposed the use of linguistic instruments to support requirements analysis. There are two main reasons for this: (i) the progress made in natural language processing, (ii) the need to provide the developers of software systems with support in the early phases of requirements definition and conceptual modelling. This paper presents the results of an online market research intended (a) to assess the economic advantages of developing a CASE tool that integrates linguistic analysis techniques for documents written in natural language, and (b) to verify the existence of potential demand for such a tool. The research included a study of the language – ranging from completely natural to highly restricted – used in documents available for requirements analysis, an important factor given that on a technological level there is a trade-off between the language used and the performance of the linguistic instruments. To determine the potential demand for such tool, some of the survey questions dealt with the adoption of development methodologies and consequently with models and support tools; other questions referred to activities deemed critical by the companies involved. Through statistical correspondence analysis of the responses, we were able to outline two "profiles" of companies that correspond to two potential market niches which are characterised by their very different approach to software development.

**Keywords:** Market research, Potential demand, NLP-based CASE tools, Requirements Analysis, Conceptual modelling

## **1. Objectives and structure of the paper**

### **Premise**

This paper presents the results of an online market research conducted in the spring and summer of 1999 by the Department of Computer and Management Sciences of Trento University, Italy. The study is part of a larger project whose principal aim is to identify the advantages and disadvantages of market research done online with respect to traditional methods and channels, and to look at its applicability in diverse product markets<sup>1</sup>. In methodological terms the objective of the research presented in this paper was to demonstrate the benefits of conducting online market studies for innovative products. Problems with such innovative products derive firstly from the fact that their

---

<sup>1</sup> Multi-year project funded by the Department of Computer and Management Sciences of Trento University.

characteristics cannot be thoroughly defined before conducting the research, and secondly their availability in commercial form usually requires further sizeable investments in research and trialling. Both of these issues are critical for CASE (Computer Aided Software Engineering) tools, which use linguistic instruments to analyse documents in natural language, and are therefore based on technologies for natural language processing (NLP) developed in the field of Artificial Intelligence. Working from the perspective of a company attempting to decide which products to develop (from among different projects related to NLP-based applications), our objective was to evaluate potential demand for NLP-based CASE tools. In conducting the study we made the reasonable assumption that the respondents (people involved in developing software systems) could be contacted easily by Internet; this prerequisite could not be guaranteed principally at a national level for other sectors studied previously (e.g., tourism or electronic commerce of groceries)<sup>2</sup>. At the same time, a certain predisposition not to participate in the study was to be expected, whether because of time constraints (noted even at the initial explorative interviews) or because of an already high level of saturation. In fact, both of these assumptions were confirmed during the course of the research. Nonetheless, we emphasise that this paper focuses on the results of the actual content of the research, and hereinafter we describe only methodological aspects that are pertinent to the interpretation of the results obtained<sup>3</sup>.

## Objectives

As previously mentioned, the aim of the research was to analyse the potential demand for a CASE tool integrating linguistic instruments as a support to requirements analysis [2]. To give the context in which such a tool could be designed and used, the following paragraph first describes the role of natural language in requirements engineering and then classifies the possible applications of linguistic instruments, making reference to the architecture of an ideal NLP system and to the three fundamental activities of requirements analysis: Elicitation, Modelling and Validation [3]. Our market research refers principally to the support of conceptual modelling, an activity that to benefit from the use of linguistic instruments requires the design of a modelling module. The other activities could be supported by existing functionalities of an NLP system, with varying levels of performance.

It was found early in the study that none of the commercial CASE tools exploited linguistic instruments to support requirements modelling [4]; this meant, therefore, that the market research was to focus on a new product whose features could not be defined in relation to similar existing products (analysis of the competition). Numerous research projects do exist in this area, however, and serve as a testimony of the considerable interest in the use of linguistic instruments in requirements engineering<sup>4</sup>. The common objective is to carry out a linguistic analysis of requirements documents in order to produce conceptual models of them<sup>5</sup>. Among the most recent projects, as an example, we can cite those described in [8,9]. While a complete review is beyond

---

<sup>2</sup> Some comparisons deriving from our research are described in [1].

<sup>3</sup> For further study of issues related to online market research, the interested reader can refer to the literature (see for example, the publications found at ESOMAR - European Society for Opinion and Marketing Research - <http://www.esomar.nl/>).

<sup>4</sup> See [5,6]. A bibliography is available at <http://nl-oops.cs.unitn.it>.

<sup>5</sup> The first proposals to use linguistic criteria for the extraction of entities and relations, and then objects and associations, from narrative descriptions of requirements date from the 1980s [7].

the scope of this paper, it is worth noting how different approaches can be analysed by looking at two principal aspects (depending on the characteristics of the linguistic tools adopted):

- a) how "natural" the input language is, which is normally subject to restrictions regarding grammar, vocabulary, or both;
- b) how much intervention by an analyst is needed in order to process "semiautomatically" the text or to identify the key elements for conceptual modelling.

The survey described in this paper focuses on the first of these points, one that we deem of vital importance because whatever the approach adopted, the "naturalness" of the language directly affects the amount of effort needed to extract useful information from the documents. First, it was necessary to establish whether the documents gathered in the requirements elicitation phase were in 'real' natural language or in some type of restricted language, and if they were in natural language, whether the user or customer could be asked to describe the requirements using a more restricted language. In fact, if the documents are written in a 'controlled' language (restrictions on grammar or vocabulary), information can be extracted using syntactic or 'shallow' techniques, such as parse trees<sup>6</sup>. To obtain equivalent performances with documents in unrestricted natural language it is necessary to have a

semantic representation of knowledge that embeds reasoning techniques. Such applications are currently being studied<sup>7</sup>. Moreover, the language used in the documents can be more or less linked to a particular application domain (for example, software for telecommunications), thus determining the degree of specialisation of the support linguistic tool to be used in the conceptual analysis, and therefore of its knowledge base. In other words, hypothesizing that the basic NLP technologies are available, for a company that must decide whether or not to invest in the development of an NLP-based tool for requirements analysis, it is important to establish first if it is possible to design and realise a general-purpose tool to support software development for different application domains or if instead it is necessary to make further investments later to customize the tool for the different companies or customers it will eventually serve. These are all essential considerations in determining the investment necessary to convert a research prototype - like those developed in the existing research projects - into a commercial tool.

Results of preliminary interviews as well as the state of the art of existing prototypes led us to decide not to investigate the degree of analyst intervention requested nor performance requested of the tool (point b: we limit ourselves on this point to giving some general findings that emerged while conducting the research). To do so would have required further investment in a more extensive market research; such study would be justifiable only with a positive outcome, certainly not guaranteed, relative to the issues related to point a). Moreover, to assess the potential market for an NLP-based tool for requirements analysis, we studied aspects related to the diffusion of methods and instruments of software engineering. In particular, we intended to verify

---

<sup>6</sup> Included in this category are, for example, the instruments described in [10] and [11].

<sup>7</sup> For example, to recognise if *Washington* is the name of a person, of an airport, or of a city in a given document requires a semantic approach. Limitations on space do not permit a deeper discussion of this issue here; see for example [12].

whether requirements analysis is in fact considered critical in relation to other important activities in software development (testing, documentation, etc.).

### **Structure of the paper**

The paper is organised as follows: the next section describes the context of an NLP-enabled CASE tool and summarises possible applications of linguistic tools for requirements engineering. This provides information on the design of the questionnaire and the eventual interpretation of the results. The third section outlines the plan of the market research, noting the different phases and focusing on the questionnaire and on the characteristics of the respondents. The main results of the online survey are presented in the fourth section, where they are analysed using a statistical technique referred to as correspondence analysis. The profiles obtained have revealed the existence of two market niches characterised by their diverse approaches to software development. Finally, some observations are given regarding the characteristics of the survey and the extendibility of the results. The conclusions summarise how the results of the survey can be used by those who develop software in general, and by those who design tools and environments for requirements analysis in particular.

## **2. The role of natural language in requirements engineering**

Much has been written on the importance of requirements analysis. In order to show why environments and tools to support such analysis are less satisfactory than those available for the other phases of the software life-cycle, we shall briefly review the distinctive features of requirements engineering, defined as:

“the systematic approach of developing requirements through an iterative cooperative process of analysing the problem, documenting the resulting observations in a variety of representation formats, and checking the accuracy of the understanding gained”. [3, p 13].

Thus evident is the central importance of communication<sup>8</sup> and knowledge. Compared with other phases of software engineering, requirements analysis and conceptual modelling [15] present unique difficulties. Many of the activities involved are cognitive and require creativity as well as knowledge about information technologies and the application domain. Moreover, the recent advances brought about by business process re-engineering (BPR) and the inclusion of innovative components in information systems are broadening the scope of projects. As a consequence, the number of the actors, interactions and languages involved have increased. Completing the picture are the needs of companies, which operate at ever higher levels of competitiveness and which demand increasingly flexible information systems.

In this context, the use of linguistic tools – more precisely of NLP systems – to support the development of software systems in general and requirements analysis in particular, may help the analyst to:

- concentrate on the problem rather than on the modelling;
- interact with other actors;
- take into account the various kinds of requirements (organisational, functional, etc.);

---

<sup>8</sup> “The hard part, and the true essence of requirements, is trying to understand your customer’s needs. A person involved in requirements needs human skills, communication skills, understanding skills, feeling skills, listening skills” [13]. See also [14].

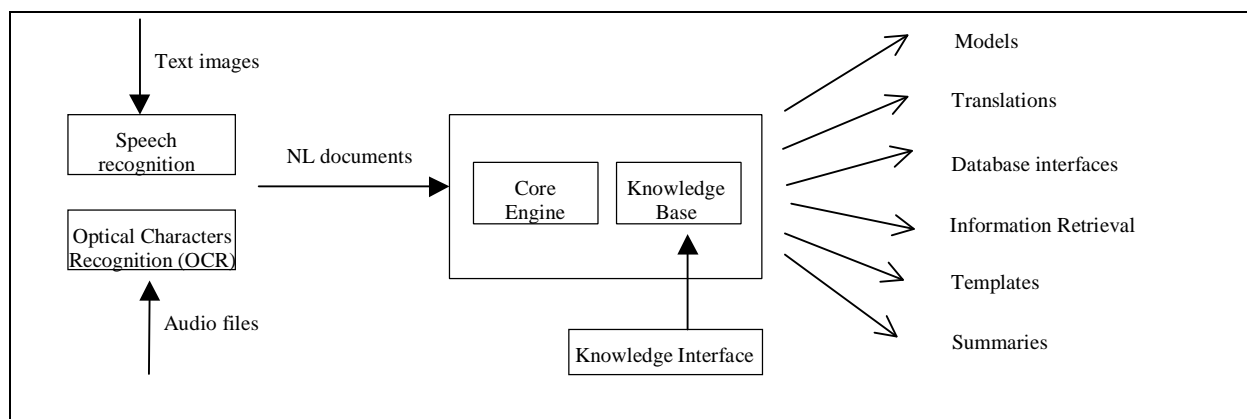
- achieve traceability as from the first documents produced;
- manage more efficiently the problem of the changing user requirements.<sup>9</sup>

As regards the possible applications of NLP systems to requirements engineering, it is worth noting that they are able to process both vocal and textual input, sometimes imposing restrictions such as limiting the vocabulary or the grammar.

NLP systems can be used to obtain, with different levels of performance, essentially three types of output:

- syntactic, semantic or pragmatic analysis;
- text either in the same language or another one, natural or artificial; -
- syntheses in the form of differently structured summaries or templates.

Figure 1 is a simplified scheme of an ideal general-purpose NLP system. It is important to remember that the systems for real applications are usually highly dependent on the task and on the domain<sup>11</sup>.



*Figure 1 – The architecture of a general-purpose NLP system*

With reference to this scheme, linguistic tools of differing complexity and especially of differing maturity can be used:

a) in the requirements elicitation phase:

- to facilitate the digitising of requirements documents using speech recognition systems or NLP-based interrogation interfaces;
- to reveal ambiguities and contradictions in documents describing user needs (see for example, [12,18,19];
- to design questionnaires or interviews, by verifying the ambiguity of the questions;
- for automatic analysis of replies to open-ended questions, interpreting and classifying their contents [20].

b) to model requirements by extracting (directly from the text) the descriptions of the elements to include in the conceptual models envisaged by the development

<sup>9</sup> For a recent study on why it is impossible for users to know their requirements beforehand, see [16]. <sup>11</sup>

On this point, see, for example, the tasks required by the MUC competitions (*Message Understanding Competition*) organised by the DARPA (Defense Advanced Research Projects Agency) [17].

method adopted, in particular UML (Unified Modelling Language)<sup>10</sup> diagrams (see Figure 2).

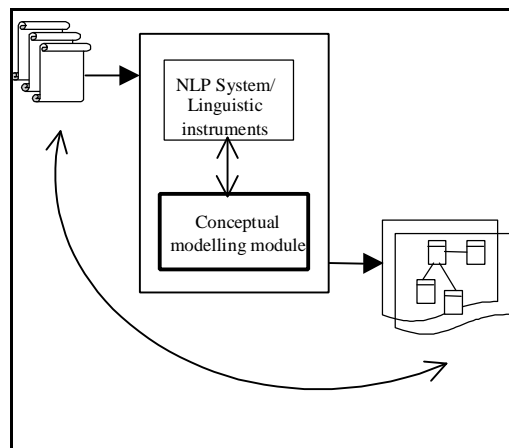


Figure 2 – The models generation process

- c) to support requirements validation, by exploiting the generation functionality of NLP systems to produce descriptions in natural language based on the structures used to represent knowledge.

A complete vision requires noting that NLP tools can also be used for documentation, generating reports on the various stages of requirements collection and modelling; for traceability, allowing a link to be maintained between the texts used and the models produced; and for the translation of documents into various languages, something that becomes increasingly necessary in the design of international information systems.

The survey described in this paper concerns the second of these points, that is, the use of NLP techniques to support the development of conceptual models, given that it requires the design of a modelling module. All the other activities could be supported by existing functionalities of an ideal NLP system, albeit with different performances. The most important assumption is that the requirements documents, once analysed, can contribute to a "knowledge base" from which to extract elements deemed useful for modelling activities. There are two important aspects to note regarding projects for developing this type of instrument: i) many of these projects are based on *ad hoc* NLP systems, and therefore do not appear to correspond to the requirements for scalability and robustness of real applications; ii) given the complexity of natural language, almost all of them expect that documents will be written in restricted language or that some revision of the text will have taken place before undergoing the automatic analysis. These two facts are worth remembering when interpreting the results of market research and when estimating potential investments in NLP technologies, and certainly when developing a CASE module to support requirements analysis.

### 3. Plan and realisation of the market research

The decision to investigate the market for an NLP-based tool for requirements analysis was taken in the context of a joint research project with the Department of Computer Sciences of Durham University (UK) in which a prototype was developed of a CASE

---

<sup>10</sup> The official documents of the UML's specifications can be found on the OMG (Object Management Group) web site: <http://www.omg.org>.

tool - called NLOOPS -,<sup>11</sup> for requirements modelling according to the object-oriented approach [21,22].

The market research described here was based on the administration of a questionnaire whose design required consideration of the experience gained throughout the development of NLOOPS, and of the methodology and techniques of online market research. Specifically, the research progressed in the following phases:

- preliminary survey
- identification of interview subjects
- designing and testing of the questionnaire
- selection of the contact method
- distribution of the questionnaire and reminders - collection and analysis of the data.

A description of each phase follows, giving greater emphasis to the third phase (designing the questionnaire) and to the final stage (analysis of data).

**Preliminary survey** The first step in the research project was to create a focus group composed of both companies that develop linguistic instruments as well as big and small businesses that develop software or offer services linked to the introduction of information technologies in the workplace. The goal of this phase was to collect information about the users' needs that could be satisfied with an NLP-based CASE tool and to gather other information useful in designing the questionnaire. The researchers were immediately confronted with pessimistic views of tools which use NLP techniques to support requirements analysis. In particular, some focus group members expressed serious doubts that the language in the documents gathered for requirements analysis was sufficiently 'natural' to justify the adoption of a tool based on NLP techniques. Others questioned the technical feasibility of such tools, citing their own unsatisfactory experiences with other NLP applications such as translation programs.

**Identification of interview subjects** In accordance with the objective of the study, the questionnaire was directed principally to persons involved in software development, and in addition to managers responsible for important decisions regarding the process of software development, including the decision to adopt methodologies and support instruments. From a statistical viewpoint, when dealing with a survey conducted via Internet, one of the main problems is to establish the degree to which the sample is representative of the target population, in this case the people or companies involved in software development. On one hand, it is reasonable to assume that the intended respondents are reachable by Internet, while on the other hand the population has characteristics (number, size, geographic distribution, etc.) that are not documented. Given this and also considering the chosen methods of contact, the approach to the study is conceptually similar to a sequential sampling. Statistically, this would classify it as a descriptive study, and as such requires caution when extending the results outside of the survey sample.

**Designing and testing of the questionnaire** Again considering the objectives of the study, in terms of both methodology and content, the survey was conducted only via

---

<sup>11</sup> Natural Language – Object-Oriented Production System, <http://nl-oops.cs.unitn.it>.



Internet and it consisted of a questionnaire on a Web page<sup>12</sup> (see appendix A). This choice was the driving force during the design and testing stage, the aim being to have a concise questionnaire with closed-ended questions in language as clear as possible.<sup>13</sup> As for the questions themselves, the choices were made as logical and pertinent issues emerged throughout the course of the focus group. After a phase of testing in which the questionnaire underwent the scrutiny - first directly and then online - of a select group of analysts and project managers, the final version was produced. The final questionnaire was divided into two sections, for a total of eighteen questions, and a final open question for further observations. The first group consisted of questions relating to the company (questions 1 – 4) and to the respondent (questions 5 and 6). The second part investigated processes of software production, so that one group of questions concerned the use of methodologies (questions 7 – 10) and tools (questions 13 and 14) in software development; another group dealt with documents used in requirements analysis (questions 11, 12 and 15) and the last three were about the efficiency of the development process (questions 16, 17 and 18). The respondents were also asked if they were interested in obtaining the results of the research or in viewing a demonstration of a prototype of an NLPbased CASE tool. The decision to introduce questions associated with an engineering approach to software development was made after verifying the possibility of using existing data. Surprisingly,<sup>14</sup> only a small amount of data was found, whether for the diffusion of object-oriented methodology or for the use of ‘classic’ models such as the entity-relationships models. These are important because the early research and conceptual models for linguistic analysis of requirements [7] looked to produce entity-relationships diagrams; moreover, these models can be seen as a particular case of the class models foreseen by the object-oriented approach. As regards the market for CASE tools,<sup>15</sup> in many cases they did not meet expectations and as a consequence did not have the desired market success [25]. We will have to wait for the adoption of the UML – developed about one year before the present research project began – as a standard for conceptual modelling by the OMG (Object Management Group); only then will there be a significant growth in the market for CASE tools, repackaged and renamed as object modelling tools or visual modelling tools. In short, the scarcity of data on the penetration and role of an engineering approach to software development influenced the choice of questions for the survey, but also, as we shall see, the ability to validate and extend the results.

The questions considered most important to verifying the existence of a market niche for an NLP-based CASE tool are those related to the documents used to collect requirements. In fact, as we have already seen, if documents are in real NL, an even more sophisticated (and costly) technology is needed to develop an environment that effectively supports analysis using linguistic instruments. It is therefore useful to establish whether the company is in a position to require clients or analysts to describe requirements in a restricted language. Typical restrictions can regard: a) grammar - aiming to have syntactic constructions that are easier to analyse by requiring, for

---

<sup>12</sup> The questionnaire is available along with the data gathered and other related research material at <http://online.cs.unitn.it>.

<sup>13</sup> For example, a questionnaire like the one used for the survey described in [23] would have to be radically altered to be used on-line.

<sup>14</sup> In light of the observations in [24], this may not be so surprising.

<sup>15</sup> The choice of tools for question 14 was made on the basis of sales data for a period prior to the study.

example, shorter phrases, using the active voice, by avoiding anaphorical references, etc.; b) vocabulary - aiming to reduce ambiguity of terms. Moreover, in order to determine the degree of customisation required of a possible NLP-based tool, further questions dealt with the level of specialisation of the terminology and the domain knowledge required to develop the software.

In the questions related to the efficiency of production processes, respondents were asked in particular about the improvements that they would like to see (choosing from a list of eight possible activities considered critical, two of which are fundamental for the phase of requirements analysis) and how they could be achieved, the choice being among 'internal delegation', 'outsourcing' and 'automation'. The final question was designed to ascertain whether the company was able to deliver the software systems or products without delays. Finally, in keeping with the general rule of market research, an incentive to participate was provided in the form of a random drawing among respondents for tickets to an opera performance at the Arena in Verona.<sup>16</sup>

**Selection of the contact method** The objectives of the research and the characteristics of the tool inherently required a contact method that would permit efficient use of time and resources while at the same time reach the largest number of potential respondents. On this point, to take into account the fact that there is a high level of saturation - due to the large number of such survey requests that the respondents receive - we had initially thought to send the questionnaire to some specialised newsgroups,<sup>17</sup> highlighting the academic nature of the research. In the first phase we identified three newsgroups whose work is related to the research topic (comp.object, comp.software-eng, alt.comp.software-tools); another twentyone newsgroups were later added to the list (the complete list is available at <http://online.cs.unitn.it>). Nonetheless, after this method of contact proved less successful than expected,<sup>18</sup> we decided to contact the companies directly by email, supplying them with the address of the Web page where they could find and complete the questionnaire. The companies' addresses were acquired online using search engines, in particular a directory of Yahoo!<sup>21</sup> (Computer > Software > Developers).

**Distributing the questionnaire and reminders** As described above, the questionnaire was administered in two different ways. In a first phase it was publicised on a number of newsgroups devoted to software development (resulting in 44 completed questionnaires and 39 software companies) and in the second, requests to take part in the survey were sent by email to 1541 addresses corresponding to 1234 software companies. By means of this second method, 107 completed questionnaires corresponding to 103 companies, were obtained. To get these results, it was necessary in many cases to send a message reminding the receiver to participate in the study, yet at the same time allowing him or her to explain the decision not to complete the questionnaire. Reasons given for not completing the questionnaire frequently referred to a lack of time and the large number of requests of this kind received (the email

---

<sup>16</sup> Because the survey concluded at the end of the Arena opera season, the tickets were replaced by CDs of opera music by Verdi.

<sup>17</sup> One of the aims of the survey, in fact, was to investigate the conditions under which newsgroups can be used to carry out online surveys.

<sup>18</sup> Limited number of questionnaires obtained (44) and accusations of spamming.

<sup>21</sup> <http://www.yahoo.com>.

messages sent are accessible online at <http://on-line.cs.unitn.it>). In addition, several addresses were incorrect, although the percentage was rather low (7.6%, 6.1% if calculated by number of companies).<sup>19</sup> Consequently, the number of valid contacts was 1424, corresponding to 1159 companies.

**Collection and analysis of the data** A total of 151 questionnaires were returned, 91% within five days of sending the initial request or the questionnaire itself. The response rate calculated for the questionnaires sent via email was around 8%. This can be regarded as a satisfactory result when compared with traditional surveys conducted by post or fax, and with other surveys of software development, for which the response rate has been 3% [25].<sup>20</sup> In strictly statistical terms, the group of companies contacted – while constituting in itself a large number - cannot be taken as a representative sample of the population of software development companies. Given this, it is important that the results be interpreted in a descriptive mode, thus requiring caution in extending them. We shall see, however, that for some questions the quality of the survey results can be evaluated by comparing them with those obtained from other surveys and with data relative to the CASE market. The results of these comparisons are provided at the end of the next paragraph.

On a methodological level, the use of newsgroups confirmed that little effort was required to ask respondents to participate, but the low number of questionnaires completed may nullify this advantage. Furthermore, the use of newsgroups should be evaluated on the basis of the following factors: level of specialisation,<sup>21</sup> number of messages, and presence of a moderator. In light of the results of our survey, in the case of very specialised newsgroups, even if the contents of the survey are relevant to them, in order to increase the response rate it is advisable to ask for the moderator's consent, or to identify one or more newsgroup leaders who can legitimate the survey with their participation.

The initial analysis noted the geographic distribution of the respondents, most of whom are residents of European states or of North America (see Figure 3). This first result of the research is supported by the analysis of similarities among different geographic distributions (using appropriate indices) showing, in fact, that these markets have similar characteristics. Given this, we present here results of the survey in its entirety, highlighting only those aspects where geographic area of residence influenced the responses.

---

<sup>19</sup> This is a rather high percentage, bearing in mind that they were collected from the homepages of official company websites. Another survey carried out in the same period on winter tourism, where the addresses were provided by a specialized magazine, found a very similar percentage of wrong addresses (8.9%), but the amount can be much higher. For example, in a survey of Internet users carried out in 1996, 35% of a total 1221 addresses were found to be wrong [26].

<sup>20</sup> This was the minimum value for the traditional-type surveys, which achieved a maximum response rate of 20%. In the survey described by Glass and Howard [25], the percentage rose to 17% after the questionnaire mailings were supplemented by telephone contacts with fax follow-up.

<sup>21</sup> For a survey on virtual supermarkets, a message was sent to 6 newsgroups obtaining 100 completed questionnaires.

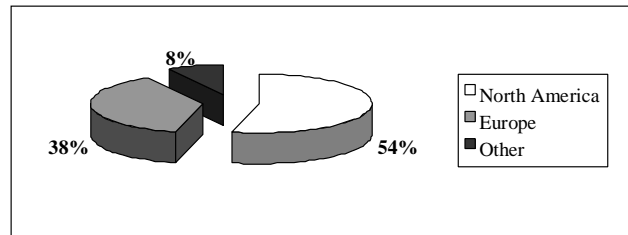


Figure 3 - The respondents by geographical area of residence

Eighty-six percent of the respondents fill roles relating to software development projects, 68% having occupied the role for more than six years.<sup>22</sup> Moreover, as to be expected, length of service influenced the position occupied in the company, so that programming work was more frequently performed by persons employed for the shortest periods, while those who had worked in their companies for 6-10 years were almost uniformly distributed among roles. To be noted is that the majority of European respondents selected 'System Engineer/Architect' but their American counterparts selected 'Project Manager', which may have been because different terms are used to denote the same role in the two areas. Some 29% of the respondents worked in companies with more than one hundred employees, although small-sized companies were also well represented (Table 1).

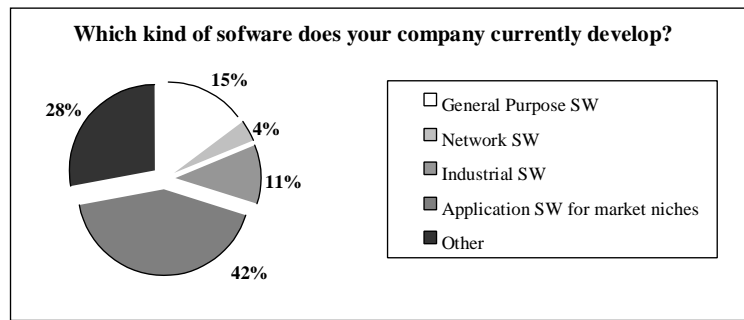
How many employees and consultants are there in your company?				
1-5	6-20	21-50	51-100	More than 100
27%	24%	15%	5%	29%

Table 1 – Company size

The core business of the companies surveyed in 77% of the cases is 'SW development' and in 23% is 'Web sites' or 'Other'. As expected, the highest percentage of companies engaged in other types of business (or rather, *also* in other types of business) consisted of larger-sized ones. As regards the type of software produced, 42% of the companies developed software for niche markets (Figure 4), with a high 48% for North America. This may be due to the presence of a larger number of small-sized companies, given that 59% of companies with five or fewer employees, and 24% of those with more than 100, operated in niche markets. Software products were mostly sold to the end-user: 84%;<sup>26</sup> only 13% sold to another software company, and 3% to software shops. Interestingly, all the companies that developed Web sites sold their products directly to the end-users, given the nature of this type of product.

<sup>22</sup> All the percentages were calculated on the total number of respondents who answered the relative questions, with non-replies omitted. <sup>26</sup>

Further investigation of this aspect would require knowledge of the number and size of the companies' customers. This, however, is beyond the scope of our survey.



*Figure 4 – Type of software*

The next paragraph provides a detailed analysis of the results of research into the existence of a potential market for an innovative tool to support conceptual analysis—a tool that has the capability to analyse documents written in varying levels of natural language.

#### **4. The results of the survey and the potential demand for an NLP-based tool to support requirements analysis**

We can identify three groups of elements that are useful in evaluating potential demand<sup>23</sup> for a CASE tool to support requirements analysis for documents written in natural language. They can be described as follows, taking into account their interrelatedness:

- [1]The market for instruments supporting software development and requirements modelling. How extensive is the market? How much competition is there? Do software developers use CASE tools? If so, which ones? (Normally the use of a CASE tool presupposes the adoption of a development methodology.) This last point was important both for establishing which conceptual models the tool should support (an aspect that became less important with the diffusion of UML<sup>24</sup>), and for reasons of compatibility and integration with existing tools<sup>25</sup>. Some information on this point could be obtained by means of the data on sales of CASE tools, but one question on this topic was inserted regarding the tools supporting requirements analysis and top-level design.
- [2]Features of the tool. The requirements principally influencing the investments necessary to develop a tool for requirements analysis based on linguistic instruments are (a) the language found in the documents gathered in the elicitation of requirements phase, crucial in identifying appropriate techniques and linguistic instruments, and (b) the degree of specialised domain knowledge required of the tool, which determines the degree of specialisation required of the producer of the CASE tool (generality). Also, given the state of the art of linguistic instruments, an

<sup>23</sup> For an introduction to the evaluation of potential demand, see for example, [27].

<sup>24</sup> In the past, the need to support different graphic notations was a drawback to the market for CASE, in that it required producers to choose which notation to support with their own tools, or to absorb the higher cost of developing different versions.

<sup>25</sup> A CASE based on linguistic techniques for object-oriented analysis does not necessarily require the realisation of an entire support environment, but rather can be seen as a module that can be integrated with an existing product.

important consideration is the performance required of the tool; in other words, how 'good' does it have to be to merit purchase?<sup>26</sup>

[3]Requirements analysis viewed as crucial. This is a vital element in identifying potential market niches and in ascertaining the propension of users to invest in a tool that supports requirements analysis, as well as their willingness and ability to accept the changes that accompany the adoption of a new tool. Companies that have an engineering approach to software development have highly standardised processes and should therefore consider the activities lacking structure or support as crucial points demanding attention. A company employing a more informal or 'craft' process would not necessarily share this concern but would, however, be more interested in the use of natural language.

To glean the most useful information on these three points, we analysed the completed questionnaires in two phases. In the first phase we looked at individual answers, studying reciprocal relationships and dependencies. In the second phase we applied correspondence analysis [28], aiming to unveil the existence of profiles corresponding to potential market niches for an innovative CASE tool.

[1] As for the use of a tool supporting requirements analysis and top-level design, only 30% replied positively. As was expected, greater use was made of these tools in large-sized companies, reaching 51% in those with more than one hundred employees, as is shown in the table of conditional distributions (Table 2). Not surprisingly, the use of these tools increases with length of service (rising from 17% to 36%) with analysts as the category of employee using them most frequently.

*Table 2 - Use of tools for requirements analysis and top-level design by company size*

Do you use any tool supporting requirements analysis and top-level design?	How many employees and consultants are there in your company?				
	1 – 5	6 – 20	21 – 50	51 – 100	More than 100
<b>Yes</b>	16%	18%	33%	33%	51%
<b>No</b>	84%	82%	67%	67%	49%

Moreover, 84% of the respondents stated that they used specific methodologies for software development. Size was a determining characteristic here, 78% of companies with five or fewer employees using specific methodologies and 93% for those with more than 100. The type of software or the sales channel does not significantly influence the use of methodologies, although role and experience seems to do so to some extent.

The best known diagrams for data modelling, entity-relationship (E-R) diagrams were used by 63% of respondents who adopted a methodology. Moreover, smaller company

<sup>26</sup> A study of the 'robustness' is of utmost importance also to establish the degree of analyst intervention required in developing requirements models, and should be conducted using a prototype of the tool. See also point (b) of the introduction and conclusion.

size corresponded to their more infrequent use (52% in companies with fewer than five employees, 73% in those with more than 100). The use of E-R diagrams was substantially greater among respondents who had worked longer in the computer business (increasing from 35% among those who had worked in the field for less than three years to 66% among those who had done so for more than ten). Finally, as regards the type of software, E-R diagrams were used to very different extents by respondents who developed general-purpose software (93%) and by those who developed network software (25%), while there were no substantial differences as far as the other items are concerned.

The percentage of respondents who used an object-oriented (OO) method was 68%, a percentage similar to that of E-R diagram users. The classification by company size shows a difference between companies with five or fewer employees (60% of which used OO methods) and those with more than 100 (74% of which do so). There are no significant variations with respect to years of experience, while there is a closer association with the position occupied within the company: the percentages ranged from 45% for programmers to 78% for system engineers/architects. An interesting comparison can be made in Table 3, where one notes that those who adopt OO methods were already accustomed to using E-R diagrams, thus indicating that they seemed more inclined to use an OO approach.

*Table 3 – Entity-Relationship diagrams and Object-Oriented Methods*

<b>Do you use an OO Method?</b>	<b>Do you use Entity-Relationship diagrams to model your data requirements?</b>	
	<b>Yes</b>	<b>No</b>
<b>Yes</b>	69%	63%
<b>No</b>	31%	37%

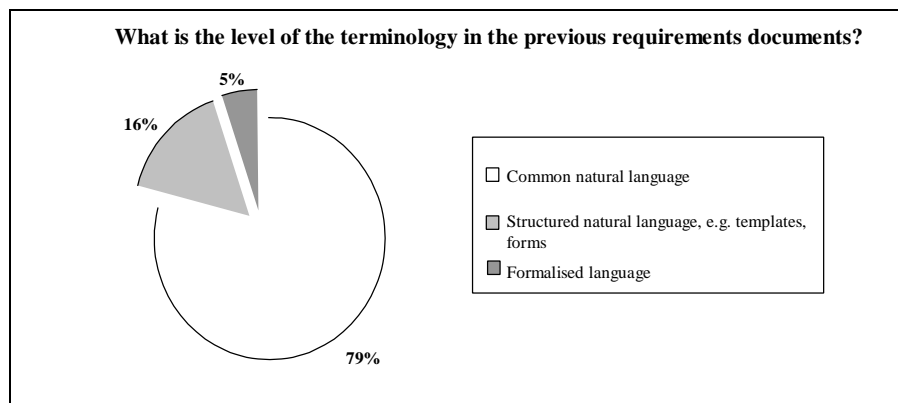
As far as the most widely used OO method, 77% of respondents who replied in the affirmative to the previous question declared that they use UML. This is a result which confirms the affirmation of UML as the industrial standard for OO modelling. It is worth mentioning that the survey was carried out approximately one and a half years after the adoption of UML by the OMG.

It also emerged that the great majority of the respondents who said that they did not use methodologies did not use tools for requirements analysis and top-level design either (90%): indeed, there is an association between the use of methodologies and CASE tools. Another finding to be emphasised is the connection between the use of CASE tools for requirements analysis or top-level design and the type of language employed in documents. Not unexpectedly, these tools were used more frequently when the language was more formal (24% with 'common natural language' and 63% with 'formalised language'). Even if these results should be treated with caution, given the low number of companies surveyed, they seemingly confirm the inability of currently available CASE tools to meet the needs of natural language processing by yielding environments that are effectively useful. As far as the tools used are concerned, 52% of respondents who replied in the affirmative to the previous question declared that

they used Rational Rose.<sup>27</sup> Rational Rose was the tool with the highest market share both worldwide and in Europe.<sup>28</sup> In 1998 it accounted for 33% of the market, with an increase of 79% on the previous year.<sup>29</sup> For this reason, the percentage found by our survey (52% for the year 1999) appears to be as one would expect.<sup>30</sup>

[2] As noted, the type of language used in requirements documents determines the complexity of the linguistic instruments and of the NLP techniques to be used. When documents are written in a constrained language (a subset of NL) – which imposes restrictions on the grammar or the vocabulary, or both – simpler and more mature linguistic tools can be used. However, it is not usually possible to impose restrictions on the language employed. Firstly, because it is necessary to adopt a customer-oriented approach in the development of software applications. Secondly, because it is necessary to reduce the risk that the restrictions imposed on the language and the formalisms adopted will force the user, or even the analyst, to express what the models permit to be represented, rather than the real requirements of the system. The survey shows that, in both Europe and North America, requirements documents are furnished directly by the customer and integrated with interviews in around two-thirds of projects. The main difference between the two regions considered was the percentage of companies that conducted interviews with customers: 73% in North America and 58% in Europe, without significant differences of behaviour between small- and large-sized companies.

With regard to the level of the terminology in requirements documents, one finds that 79% of the latter are couched in natural language (Figure 5). For the correspondence analysis, the final two modalities (structured and formalised language) have been merged.



*Figure 5 – Level of terminology in the requirements documents*

<sup>27</sup> None of the tools indicated by those choosing the option 'Other' was selected more than twice.

<sup>28</sup> International Data Corporation (IDC) data.

<sup>29</sup> These figures seem to contradict the results of the survey by Glass and Howard [25], where CASE technologies are described as being in decline. However, it should be pointed out that where back-end or 'lower' CASE are concerned, many of the functions offered by these tools are by now part of the development environment. Moreover, other expressions are often used instead of 'CASE': for example, the IDC surveys use OOAMDC (Object-Oriented Analysis, Modelling, Design and Construction) tools. On the other hand, in 1998 the market for OOAMDC grew by more than 10% (24% in Europe), See also the results in [29].

<sup>30</sup> It should be pointed out, however, that the data of our survey are expressed in terms of units of output by the companies surveyed, while the sales figures are calculated on invoices and consequently depend on the prices charged by vendors.



An analysis of the interdependence of the use of natural language with the other factors examined did not show any significant association with type of company, nor with the adoption of a methodology.

Another important aspect concerning both the potential demand for an NLP-based CASE tool in particular and software development in general, is the domain knowledge required for adequate understanding of the problem so that the user's requirements can be defined. In fact, in the presence of high levels of specialist knowledge, the tool must be adapted to the needs of every customer if it is to operate efficiently in different corporate settings. By contrast, a very low level permits the development of a single standard tool able to operate in different fields of application. In this regard, it was found that respondents required an average (54%) to high (34%) level of domain knowledge. It also emerged, that the higher the level of domain knowledge required to develop the software, the greater the use of methodologies (9% for low levels, 53% for average ones, and 38% for high ones) and of tools for requirements analysis and top-level design (2%, 56% and 42% respectively).

[3] As regards the efficiency of production processes, upon conclusion of the market study it was important to determine which software activities were viewed as crucial, as well as their weight relative to requirements (question 16).

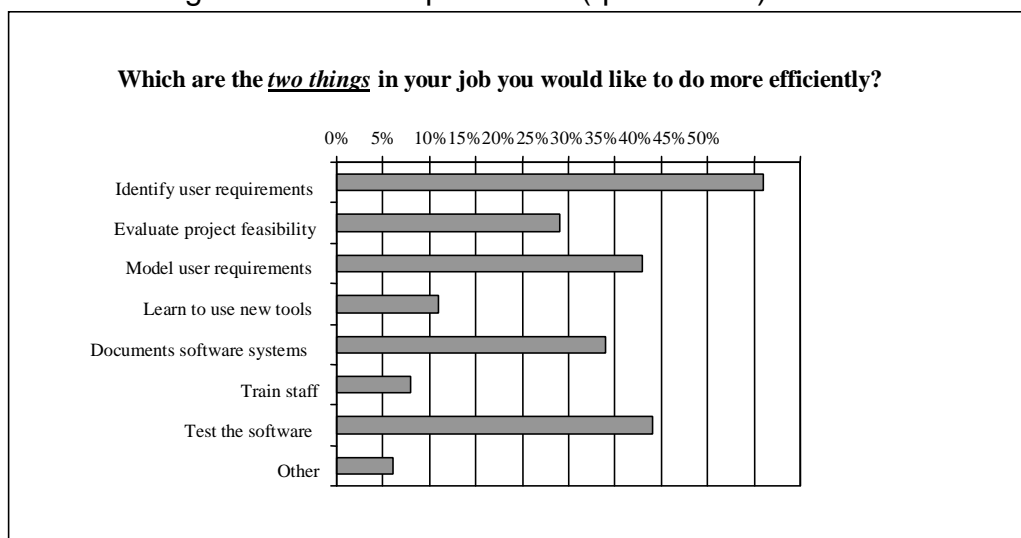


Figure 6 – Activities perceived as crucial in software development

In interpreting the answers to this question, it is worth noting that two selections were requested, thus having results above 100 percent. Figure 6 shows that 'Identify user requirements' and 'Model user requirements' were cited as priorities by a high percentage of respondents.<sup>31</sup> Unlike in the case of 'Identify user requirements' – which was largely independent of the language used to model requirements (46% for common natural language, 37% for structured natural language and 50% for formalised

<sup>31</sup> To be noted is that also around one-third of the final observations concerned the role and importance of requirements. Taking into account of the different goals of the surveys described in [30,31], we can compare these results with those obtained for a question therein on the perceived relative importance of software problems in Europe (most of the software problems are in the area of requirements specification and managing customer requirements; following documentation and testing) and on the perceived scope of a generic process model (defining system requirements, 78%).

language) and for 'Testing the software' (35%, 32%, 38% respectively) – for 'Model user requirements' the percentages were 38% for common natural language and 13% for formalised language, in accordance with expectations. Another noteworthy finding is that testing was viewed as crucial by higher percentages (ranging from 19% to 46%) of the respondents who used no tools at all. A similar pattern is displayed by the level of domain knowledge necessary, where at low levels of knowledge, testing was perceived as more important than all the other activities (63%, compared to 32% and 30% for medium to high levels of knowledge). Also of interest is the fact that 'Learn to use a new tool' was selected by a higher percentage of respondents declaring that they did not use a tool for requirements analysis than by those who instead said that they used a tool of this kind.<sup>32</sup>

The importance of this question requires a comparison of the results for Europe and North America (see Figure 7). Also the correspondence analysis - reported in the second part of this section - was done taking into account the centrality of this question with respect to the objectives of the market research, in which the activities considered most critical become determinative when identifying profiles.

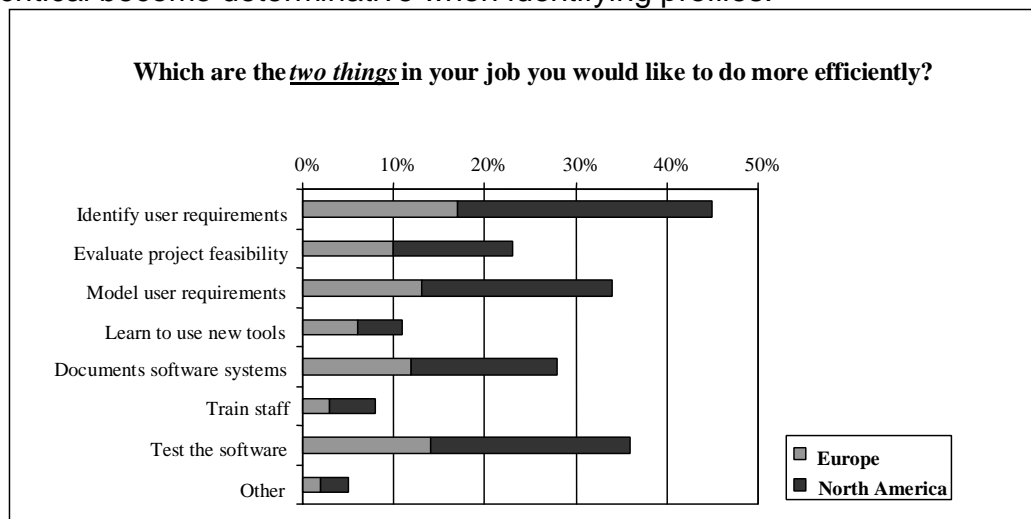


Figure 7 – Activities perceived as crucial in software development (Europe vs. North America)

To the question 'What would be the most useful thing to improve general day-to-day efficiency?', the majority (64%) chose the option 'automation', while 'outsourcing' was selected by 7% and 'internal delegation' by 29%. Contrary to expectations, no particular differences emerged among the replies to this question with respect to company size, where the only significant difference concerned companies with 6 to 20 employees, where the percentage selecting 'internal delegation' was nearly double that for other company groups, a difference which may be due to organisational shortcomings. Interestingly, the percentage of respondents who used a methodology or a requirements analysis tool and believed it less important to increase the level of internal delegation was above the average of the entire sample. Instead, there were no differences regarding the documents available for requirements analysis.

<sup>32</sup> In this regard we quote a remark made in one of the questionnaires: "I hate to be a cynic, but there are hardly any worthwhile tools. The overhead in learning to use them is too great for the payoff".

Joint analysis of the two questions on the efficiency of software production processes shows that a larger percentage of respondents who believed it important to increase the level of automation had previously selected 'Learn to use a new tool' and 'Model user requirements' (Table 4).

*Table 4 – Efficiency of software development processes*

<b>Which are the two things in your job you would like to do more efficiently?</b>	<b>What would be the most useful thing to improve general day-to-day efficiency?</b>		
	<b>Automation</b>	<b>Outsourcing</b>	<b>Internal delegation</b>
<b>Identify user requirements</b>	69%	9%	22%
<b>Evaluate project feasibility</b>	44%	12%	44%
<b>Model users requirements</b>	<b>75%</b>	4%	21%
<b>Learn to use new tool</b>	<b>86%</b>	0%	14%
<b>Documents software systems</b>	71%	5%	24%
<b>Train staff</b>	18%	0%	82%
<b>Test the software</b>	67%	4%	29%
<b>Other</b>	43%	14%	43%

For the final question, regarding the average delay in delivery of the software, the best performances were achieved by companies with 6-20 employees (29% of which delivered with less than one week of delay and 59% with less than one month) and by those who sold directly to the end-consumer (probably for contractual reasons). Though not to a statistically significant extent, companies using formalised language delivered with the least delay, although there were no substantial differences as regards delays of more than one month (26% for common natural language, 33% for structured natural language, 25% for formalised language). A fair interpretation of these results requires one to remember that the answers do not factor in the length of the projects. Nonetheless, assuming that an average delay of less than one week corresponds to companies which on average deliver the software within the designated time, similar findings are reported in [32], where more than 80% of the respondents stated that their projects were sometimes or usually late.

Considering the purpose of this study, and particularly the question of whether there is a market for an NLP-based CASE tool for requirements analysis, the results presented thus far confirm the perception of requirements analysis as crucial for the development of systems, the widespread use of the object-oriented approach and of UML, and the important role of natural language. Specifically:

- More than 80% of the companies adopt a methodology to develop their software, and nearly 68% of them adopt an object-oriented method (UML or one of the methods merged into UML).
- The majority of the documents available for requirements analysis are in natural language and are either furnished by the customer or obtained by means of interviews.

- The domain knowledge required is medium to high.
- Tools supporting requirements analysis and top-level design are used in less than one-third of cases.
- However, identifying and modelling requirements are perceived as being at least as important as testing the software.
- A higher level of automation is indicated by around 64% of the respondents as the most useful means to improve day-to-day efficiency.

All of these elements work together to confirm the existence of a potential demand for a

CASE tool based on NLP. To justify this claim, we undertook a correspondence analysis (CA) study. This meant using a statistical technique suited for the study of relationships between modalities with two or more distinguishable variables, usually qualitative. The main steps of correspondence analysis are concisely described as follows:

- 1) define a cloud of points (rows and columns of a contingency table) in a multidimensional vector space;
- 2) choose the metric structure on this space;
- 3) produce the fit of the cloud in 1) to a variable low-dimensional subspace onto which the points (row and column profiles) are projected for display;
- 4) give an interpretation of the clusters of points corresponding to the projections of the rows and columns of the original contingency table; analyse their absolute contributions as guides to the interpretation of the underlying dimensions and their relative contributions (the so-called squared correlations) to indicate how well the points are described along the considered dimension.

The geometry of CA is very similar to Karl Pearson's [33] geometric description of Principal Components Analysis. The closeness of the points to a line, plane, or in general to a lowdimensional subspace, is defined as the sum of squared distances from the points to the subspace. In general, it is important to avoid the direct comparison of the distances among the projections of row and column profiles because they belong to different low dimensional subspaces and the raw interpretation of their distances may produce misleading conclusions.

Here we have considered a CA involving one of the items of the questionnaire (what should be done more efficiently) as dependent variable and some other collected variables (number of employees, core business, kind of software produced, use of any methodology, starting documentation, level of terminology, use of any tool, knowledge of domain, thing to improve the day-to-day efficiency, average delay in delivering the software) as independent variables in order to verify whether and how much the answer to this item is influenced by the modalities of the other variables and to identify some relevant aggregations of modalities which can reveal the potential market demand for a CASE tool based on NLP.

We present here the result of the application of the CA based on the responses to the question regarding which activities are considered most critical (see Figure 8).<sup>33</sup>

An initial interpretation of the graph can be reached by looking at the axes. Specifically, one can interpret the vertical axis in organisational terms, assuming that the request for more automation rather than internal delegation is due to an already more or less solid organisational structure. The horizontal axis, meanwhile, corresponds to an engineering or to a more informal approach to software development depending on the use or not of methodologies and instruments to support analysis and designing.

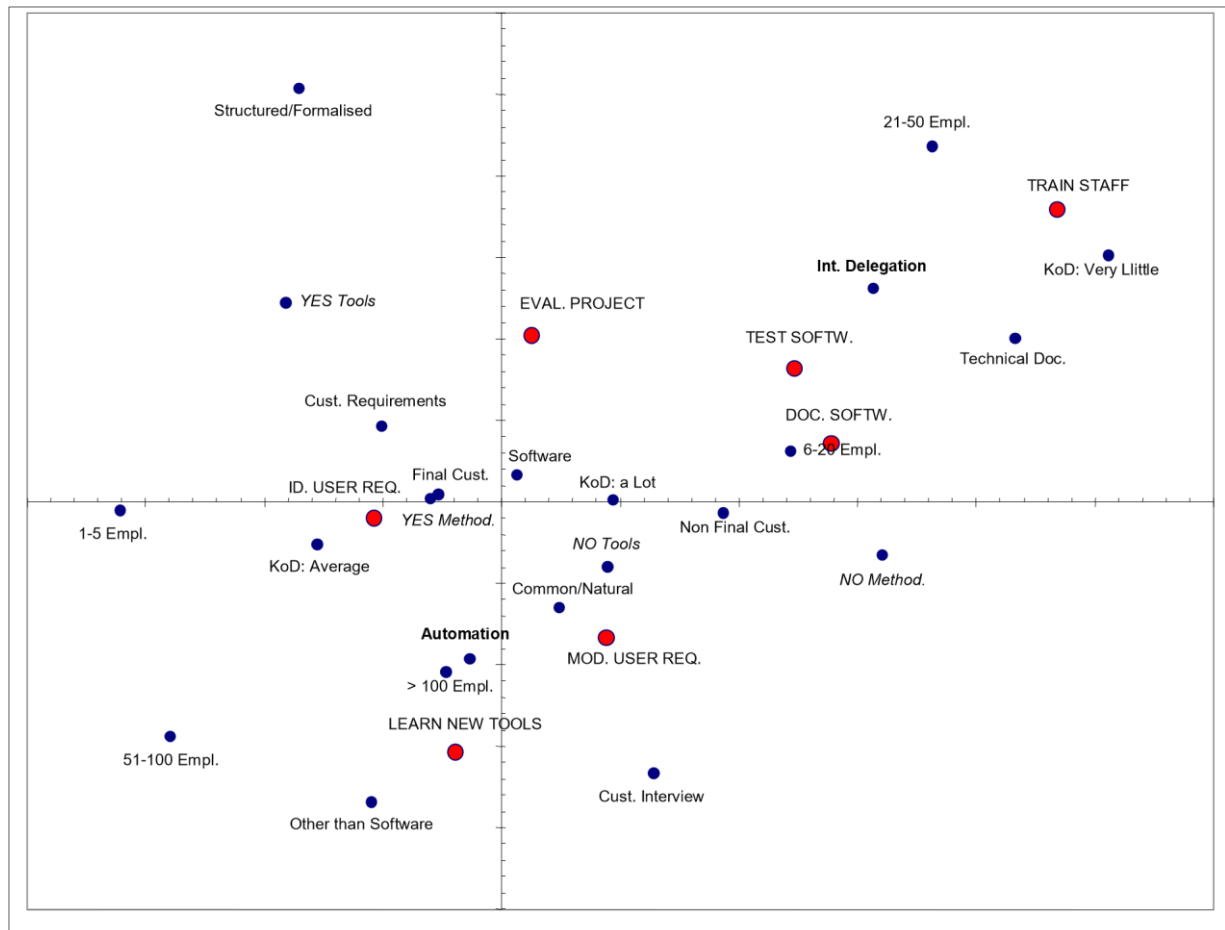


Figure 8 - Output of the correspondence analysis<sup>34</sup>

According to this interpretation of the graph, there are two potential market niches.

The first market niche corresponds to companies that adopt methodologies and instruments to support requirements analysis and top-level design. We can safely assume that they use an 'industrial' rather than 'craft' software development process. For this type of company, project evaluation is considered a critical activity, along with requirements identification. These two activities, among the possible activities listed in the questionnaire, are the most interdisciplinary and at the same time the most difficult

<sup>33</sup> The contingency table is available at <http://on-line.cs.unitn.it>.

<sup>34</sup> Two points ('Other' in question 16 regarding critical activities, and 'Outsourcing' for question 17) have not been represented because of their great distance from the centre (low frequency), thereby making the graph more comprehensible.

to structure. In particular, for purposes of our study, requirements identification can be efficiently supported by tools able to analyse documents in natural language. Moreover, for this type of company, the tool should be specialised to have an appropriate level of domain knowledge for the given area of software development. The client provides requirements documents and the software produced is in turn delivered to the client. For a customer-oriented approach, this means having only a limited possibility to ask the client to write the documents in a restricted form of natural language; however, these companies sometimes receive the documents in a somewhat structured (formalised) form. In these cases it is possible to envision the use of less sophisticated linguistic techniques to analyse requirements documents in order to produce conceptual models using the object-oriented approach.

The second market niche includes medium- or large-sized companies that use neither methodologies nor instruments to support requirements analysis and top-level design. They do, however, perceive requirements modelling as critical, along with other activities such as software documentation and testing, which are already supported in varying ways by existing CASE tools. One can reasonably conclude that also this second group of companies constitutes a market niche for a CASE tool enabled by linguistic instruments. In fact, a CASE of this type could integrate the functionalities of a traditional CASE, favouring the adoption of an engineering approach in software development. Another activity deemed critical is to learn new tools, an obstacle that could be surmounted by adopting a CASE that makes extensive use of natural language. The indication of requirements modelling rather than identification brings to light the fact that a problem at the level of requirements specification can hide deeper problems related to requirements elicitation (these can be supported by speech recognition systems and by all the functionalities envisaged in point a of section 2.). This is confirmed to some extent by the fact that identification, rather than modelling, of requirements is considered critical by the companies that adopt a more structured approach to software development.

An important aspect of this research is the broader application of the results. As noted, this research is descriptive, based on a large number of questionnaires (among the highest we have seen in our studies<sup>35</sup>), yet not fully representative of the population. The fact is that for the software industry, there simply is not enough information on the reference population to permit a meaningful and statistically correct extension of the results.

Having said this, we maintain that it is useful to make a comparison with data available in the literature. The following table summarises the most significant of these (Table 5). Worth noting is the scarcity of existing data. Although the surveys to which these results refer<sup>36,37</sup> are very different, their similarities do stand out.

*Table 5 – Comparison with results relative to other surveys and the CASE market*

---

<sup>35</sup> Notable exceptions are the surveys conducted by the European Software Institute: <http://www.esi.es>.

<sup>36</sup> These surveys were carried out with different objectives and using different methods and samples. The survey described in [25] used 78 questionnaires compiled mainly by directors or managers of information systems development in companies operating outside the software field, while the Finnish one reports results relative to 12 Finnish companies, 8 of which worked exclusively in the software field.

<sup>37</sup> See [23]. 42

See [25]. Note that when this survey was carried out, UML had only just been adopted as standard by OMG.

	<b>NLP-based CASE tool online Market Research – 1999 (142 companies)</b>	<b>State of the practice Survey on RE - 1999<sup>41</sup> (12 companies)</b>	<b>SW Development - State-of-the Practice - 1997<sup>42</sup> (78 companies)</b>	<b><i>Market share OO CASE tools – 1998<sup>43</sup></i></b>
<b>Sell to the end-user</b>	<b>84%</b>	83%	-	
<b>SW as core business</b>	<b>82%</b>	66%	-	
<b>Use OO methods</b>	<b>68%</b>	50%	39% (Use O-orientation)	
<b>Use UML</b>	<b>77%</b>		53% (Use a formal life cycle methodology)	
<b>Natural language requirements</b>	<b>79%</b>			
<b>Use RA tools</b>	<b>30%</b> (& top level design tools)	-	-	<b>&gt; 48%</b>
<b>Use Rational Rose</b>	<b>52%</b>	100%	-	
<b>Identify user requirements</b>	<b>46%</b>	0%	29% (Use front-end CASE tool)	
		-	-	<b>33%</b>
		66%	-	

We can also cite here some data found in [34], which contains detailed indications of the percentage of pages in natural language or similar forms – text with keywords, hierarchical enumeration and table – for three projects, having values ranging from 82% to 99% (73%, 43.9% and 34.4% respectively, only for natural language text).

Another aspect that enables positive assessment of the outcome of the survey is the low percentage of non-replies (1.65%) and the fact that in the case of replies for which the option 'other' was selected, in 91% of cases a specification was given.

## 5. Conclusions

As the principal aim of this research project was to assess if there is a market for NLP-enabled CASE tools, the most important finding is that the majority of the documents available for requirements analysis are provided by the customer and couched in 'real' natural language, leading to the conclusion that the use of linguistic techniques and tools may perform a crucial role in providing support for requirements analysis.

Because an engineering approach suggests the use of linguistic tools suited to the language employed in the narrative description of user requirements, we find that in a majority of cases it is necessary to use NLP systems capable of analysing documents in full natural language. If the language used in the documents is controlled (giving a

subset of natural language), it is possible to use simpler and therefore less costly linguistic tools, which in some cases are already available. Instruments of this type can also be used to analyse documents in full natural language, even if in this case more analyst consultation is required to reduce the complexity of the language used in input documents or to intervene automatically in the models produced as output. Moreover, needed in many cases, besides an adequate representation of the shared/common knowledge, is specialised knowledge of the domain. Once again, the management of expert knowledge requires more substantial investments to adapt the tool to the company's needs.

As for the potential demand for NLP-based CASE tools, two company profiles have been identified, corresponding to two distinct market niches. The first is composed of companies having an engineering approach to software development and that indicated - of the two activities linked to requirements analysis - the identification of requirements as the more critical. In this case the tool could be configured as a module to integrate with the CASE tool already used by the company, and would provide support for phases where existing tools are insufficient. In the second market niche, the technologies of natural language are used to facilitate the adoption of a CASE tool and more generally of 'best practises' of software development, given that along with requirements modelling, these companies have also indicated as crucial activities in which the contribution of software engineering is well developed (testing or software documentation, for example).

We can also make some preliminary observations here regarding the features expected of a tool based on NLP, proceeding from interviews with systems analysts/engineers and project managers in both small- and medium-sized companies. Specifically, they confirm assumptions made regarding potential demand and interest in the following features:

- The possibility to accelerate the production of analysis models and to rapidly create models to be used in interactions with users and in project groups. The fact that, for example, the class models may contain spurious classes or that some classes may be missing was regarded as less important if the models are produced automatically.

<sup>43</sup> International Data Corporation (IDC) data.

- The tool was also regarded as useful for the training of analysts, with the presentation of texts and the corresponding models, both for junior analysts and for the retraining of those unfamiliar with the object-oriented approach (the latter problem seems to be more important for small-sized companies).
- The possibility of integrating the tool with CASE tools for drawing diagrams using the elements singled out by the algorithm and using tools for documents management.

Finally, for some questions in the survey (e.g., the use of methodologies and E-R models, the use of support tools in the initial phases of development) the contributions this paper makes to the field go beyond the confines of the market research as described by the title. It confirmed some expectations (the diffusion of the object-oriented approach), which on the surface could appear obvious, yet have not been



sufficiently supported by hard data. It also confirmed the presence of significant possibilities for the adoption of instruments and methods of software engineering [35].

## References

- [1] Franch M, Mich L, Osti L. Online Research as Decision Tool for Marketing and Management Strategies. In Proc. Information Technology for Business Management - ITBM2000, 16th IFIP WCC, Beijing, China, 21-25 August 2000, Gan R (ed), Beijing, 2000, pp 737-743.
- [2] D'Elia M. On-line Market Research: an Application to the Software Domain. Degree Thesis, University of Trento (In Italian), 2000.
- [3] Loucopoulos P, Karakostas V. System Requirements Engineering. McGraw-Hill 1995 [4] Chiocchetti N, Mich L. The Market for Object-Oriented CASE tools. Tech Report, Department of Computer and Management Sciences, University of Trento (In Italian), 31, 2000.
- [5] Burg J.F.M. Linguistic Instrument in Requirements Engineering, IOS, Amsterdam, 1997. [6] Ryan K. The Role of Natural Language in Requirements Engineering. IEEE 1992; 240242.
- [7] Chen PP-S. English Sentence Structure and Entity-Relationships diagrams. Information Sciences, 1983; 29: 127-149.
- [8] Ambriola V, Gervasi V. An Environment for cooperative construction of natural language requirements bases. In Proc. 8<sup>th</sup> ICRE, IEEE Computer Society Press, 1999, pp 124-130.
- [9] Juristo N, Moreno AM, Lòpez M. How to use Linguistic Instruments for OO Analysis. IEEE SW, May/June 2000, 80-89.
- [10] Fuchs NE, Schwitter R. Attempto Controlled English. In: CLAW'96, 1st Int Workshop on Controlled Language Applications, Katholieke Universiteit, Leuven - Belgium 1996.
- [11] Delisle S, Barker K, Biskri I. Object-Oriented Analysis: Getting Help from Robust Computational Linguistic Tools. In G Friedl, HC Mayr (eds) Application of Natural Language to Information Systems, OCG, pp 167-172, 1999.
- [12] Mich L, Garigliano R. Ambiguity Measures in Requirements Engineering. In Proc. ICS2000 16th IFIP WCC, Beijing, China, 21-25 August 2000, pp 39-48.
- [13] Davis AM. The Harmony in Rechoirments. IEEE Software, March/April 1998, pp 6-8.
- [14] Nitto E Di, Fuggetta A. Change vs Consolidation: a Challenge for SW development organisations. Rivista di Informatica, AICA 1995, 25 (4): 267-279.
- [15] Mylopoulos J. Information Modeling in the Time of the Revolution, Information Systems, May 1998; 23(3-4): 127-156.
- [16] Rugg G, Hooper S. Knowing the unknowable: the Causes and Nature of Changing Requirements. In J Eder, N Maiden, M Missikoff (eds) Proc. 1st Int. Workshop EMRPS'99, Venice, September 25-27, 1999, pp 183-192.
- [17] AAA Proceedings Message Understanding Conference. MUC-3, MUC-4, MUC-5, MUC-6, MUC-7 Morgan Kaufmann 1991, 1992, 1993, 1995, 1998, [http://www.itl.nist.gov/iaui/894.02/related\\_projects/muc/index.html](http://www.itl.nist.gov/iaui/894.02/related_projects/muc/index.html).
- [18] Fabbrini F, Fusani M, Gervasi V, Gnesi S, Ruggieri S. Achieving Quality in Natural Language Requirements. In: Proc Int SW Quality Week, S.Francisco CA, May 1998

- [19] Laitenberg O, Atkinson C, Schlich M, El Emam K. An Experimental Comparison of Reading Techniques for defect detection in UML design documents. J of S&SW, North Holland-Elsevier 2000; 53: 183-204.
- [20] Canzano G. Natural Language Processing in Market Research: Automatic Analysis of Replies to open-ended Questions. Degree Thesis, University of Trento (In Italian), 1999.
- [21] Mich L. NL-OOPS: From Natural Language to OO Requirements using the Natural Language Processing System LOLITA. In: J of Natural Language Engineering, Cambridge University Press 1996; 2 (2): 161-187.
- [22] Mich L, Garigliano R. The NL-OOPS project: OO modelling using the NLPS LOLITA. In Proc. 4<sup>th</sup> Int. Conf. NLDB'99, Klagenfurt, June 17-19 1999, Friedl G, Mayr HC (eds), Application of Natural Language to Information Systems, Wien 1999, 215-218.
- [23] Nikula U, Sajaniemi J, Kaelviaeinen H. A State-of-the-Practice survey on Requirements Engineering in Small- and Medium-Sized Enterprises. Research Report 1. Lappeenranta University of Technology, 2000.
- [24] Zvegintzov N. Frequently Begged Questions and how to answer them. IEEE SW March/April 1998; 93-96.
- [25] Glass R, Howard A. Software Development State-of-the-Practice. Managing System Development, June 1998, 7-8.
- [26] Comley P. The Use of the Internet as a Data Collection Method. SGA Market Research, 1996.
- [27] Wheelwright SC, Makridakis S. Forecasting Methods. John-Wiley & Sons, New York 1985.
- [28] Greenacre JM. Theory and Application of Correspondence Analysis. Academic Press, New York 1984.
- [29] Dutta S, Lee M, Wassenhove L Van. Software Engineering in Europe: A Study of Best Practices. IEEE SW May-June, 82-90, 1999.
- [30] ESI, ESPITI - European User Survey Analysis. European Software Insitute, Spain, Nov 1996.
- [31] ESI, System Engineering in Europe. Survey: Summary of Results. European Software Insitute, Spain, Aug 1998.
- [32] Genuchten M van. Why is Software Late? An Empirical Study of Reasons for Delay in Software development. IEEE Trans on SWE, June 1991; 17 (6): 582-590.
- [33] Pearson K. On Lines and Planes of closest Fit to Systems of Points in Space. Philosophical Magazine, 1901, ser. 6 (2): 559-572.
- [34] Melchisedech R. Investigation of Requirements Documents Written In natural Language, Requirements engineering. Springer-Verlag 1998, 3:91-97.
- [35] ESI, Software Best Practice Questionnaire, Analysis of Results. European Software Insitute, Spain, Dec 1997.

## **Figures**

Figure 1 - The architecture of a general-purpose NLP system

Figure 2 - The models generation process

Figure 3 - The respondents by geographical area of residence

Figure 4 - Type of software

Figure 5 - Level of terminology in the requirements documents

Figure 6 - Activities perceived as crucial in software development

Figure 7 - Activities perceived as crucial in software development (Europe vs. North America)

Figure 8 - Output of the correspondence analysis

## **Tables**

Table 1 - Company size

Table 2 - Use of tools for requirements analysis and top level design by company size

Table 3 - Entity-Relationship diagrams and Object-Oriented Methods

Table 4 - Efficiency of software development processes

Table 5 - Comparison with results relative to other surveys and the CASE market

## **On-line material (<http://on-line.cs.unitn.it>)**

Questionnaire (html form)

Contacted newsgroups list

E-mail messages

Correspondence analysis

## **Appendix A Questionnaire for a new CASE tool**

1. How many employees and consultants are in your company?

- 1-5
- 6-20
- 21-50
- 51-100
- More than 100

2. What is the core core of your company?

Software

- Web sites
- Other: \_\_\_\_

3. What kind of programs are you currently developing in Quseir?

- General purpose software

Network software

- Industrial programs
- Application program for market outlets
- Other: \_\_\_\_

4. Do you usually sell their products...

- To the end customer
- To a software company
- To offices

Other

- What is your current dominant role in the company?

- Analyst
- designer
- Programmer
- System Engineer / Architect
- project manager
- Other: \_\_\_\_

## جمع المتطلبات

يشير جمع المتطلبات إلى عملية جمع وتوثيق الاحتياجات والأهداف والقيود المحددة لأصحاب المصلحة لمشروع أو منتج. تستخدم هذه المعلومات لتوجيه تصميم الحل النهائي وتطويره واختباره. يتضمن جمع المتطلبات الفعال التعاون مع أصحاب المصلحة ، والتواصل الواضح ، والفهم الشامل للمشكلة التي يتم حلها.

### • فيما يلي خطوات جمع المتطلبات بشكل فعال:

- (1) تحديد أصحاب المصلحة: تحديد من سيتأثر بالمشروع ومن يحتاج إلى تقديم المدخلات.
- (2) المقابلات: اجتماعات فردية أو جماعية مع أصحاب المصلحة لجمع المعلومات وفهم احتياجاتهم.
- (3) الدراسات الاستقصائية: استبيان مكتوب يوزع على أصحاب المصلحة لجمع المعلومات والتغذية الراجعة.
- (4) ورش العمل: جلسات تعاونية مع أصحاب المصلحة لتحديد المتطلبات وتحديد أولوياتها والتحقق من صحتها.
- (5) الملاحظة: مراقبة وتسجيل أنشطة أصحاب المصلحة لجمع المعلومات حول إجراءات عملهم ومتطلباتهم.
- (6) النماذج الأولية: إنشاء نموذج وظيفي للحل لجمع التعليقات والمتطلبات من أصحاب المصلحة.
- (7) قصص المستخدم: تقنية سردية لوصف الوظيفة المطلوبة وتجربة المستخدم للحل.
- (8) مراجعة المستندات الموجودة: فحص الوثائق الموجودة مثل العمليات التجارية وأدلة المستخدم وتحليل المنافسين لجمع المتطلبات.
- (9) مجموعات التركيز: الجمع بين مجموعة صغيرة ومتنوعة من أصحاب المصلحة لمناقشة وتقديم مدخلات بشأن المتطلبات.
- (10) إمكانية تتبع المتطلبات: تتبع العلاقة بين المتطلبات ومخرجات المشروع لضمان الوفاء بها.
- (11) تحليل أصحاب المصلحة: تحليل اهتمامات وتأثيرات وتوقعات أصحاب المصلحة لجمع المتطلبات.
- (12) نمذجة العمليات التجارية: تمثيل مرئي للعمليات التجارية لجمع المتطلبات وتحسين الكفاءة.
- (13) الهندسة العكسية: فحص منتج أو حل حالي لتحديد متطلبات حل جديد.
- (14) جلسات JAD (تطوير التطبيقات المشتركة): جلسات تعاونية مع أصحاب المصلحة والخبراء التقنيين لجمع المتطلبات.
- (15) رسم الخرائط الذهنية: أداة مرئية لتنظيم وتصنيف المعلومات والمتطلبات.
- (16) تحليل السبب الجذري: تقنية لحل المشكلات تساعد في تحديد الأسباب الكامنة وراء المشكلة لجمع المتطلبات.
- (17) البحث الإثنوغرافي: طريقة بحث تتضمن مراقبة المستخدمين والتفاعل معهم في بيئتهم الطبيعية لجمع المتطلبات.
- (18) نمذجة المفاهيم: تمثيل مرئي للمفاهيم والعلاقات الأساسية التي ينطوي عليها المشروع لجمع المتطلبات.
- (19) اختبار الأداء: اختبار أداء الحل في ظل ظروف مختلفة لجمع متطلبات تحسين الأداء.

- (20) التحليل التنافسي: فحص عروض ونقاط قوة المنافسين لجمع متطلبات حل فريد.
- (21) تحليل الفجوة: مقارنة الحالة الحالية للنظام مع حالته المستقبلية المرغوبة لتحديد وجمع متطلبات التحسين.
- (22) تحديد أولويات المتطلبات: تحديد أولويات المتطلبات بناء على أهميتها وتأثيرها على المشروع.
- (23) تحليل التكلفة والفائدة: تقنية تقوم بتقييم تكاليف وفوائد المتطلبات لاتخاذ قرارات مستنيرة بشأن المتطلبات التي يجب متابعتها.

## خلاصة

اقترحت العديد من الدراسات في الأشهر الأخيرة استخدام الأدوات اللغوية لدعم تحليل المتطلبات. هناك سببان رئيسيان لذلك: (i) التقدم المحرز في معالجة اللغة الطبيعية ، (ii) الحاجة إلى تزويد مطوري أنظمة البرمجيات بالدعم في المراحل المبكرة من تعريف المتطلبات والنمذجة المفاهيمية. تقدم هذه الورقة نتائج بحث السوق عبر الإنترنت الذي يهدف إلى (أ) تقييم المزايا الاقتصادية لتطوير أداة CASE التي تدمج تقنيات التحليل اللغوي للوثائق المكتوبة باللغة الطبيعية ، و (ب) التحقق من وجود طلب محتمل على هذه الأداة. تضمن البحث دراسة اللغة - التي تتراوح من طبيعية تماما إلى مقيدة للغاية - المستخدمة في الوثائق المتاحة لتحليل المتطلبات ، وهو عامل مهم بالنظر إلى أنه على المستوى التكنولوجي هناك مفاضلة بين اللغة المستخدمة وأداء الأدوات اللغوية. ولتحديد الطلب المحتمل على هذه الأداة، تناولت بعض أسئلة الدراسة الاستقصائية اعتماد منهجيات التنمية، وبالتالي النماذج وأدوات الدعم؛ وأشارت أسئلة أخرى إلى الأنشطة التي تعتبرها الشركات المعنية حاسمة. من خلال تحليل المراسلات الإحصائية للردود ، تمكنا من تحديد "ملفين شخصيين" للشركات التي تتوافق مع اثنين من منافذ السوق المحتملة التي تتميز بنهجها المختلف تماما لتطوير البرمجيات.

**الكلمات الأساسية:** أبحاث السوق ، الطلب المحتمل ، أدوات CASE القائمة على البرمجة اللغوية العصبية ، تحليل المتطلبات ، النمذجة المفاهيمية .

## 1. أهداف وهيكل الورقة

### فرضيه

تقدم هذه الورقة نتائج بحث السوق عبر الإنترنت الذي أجري في ربيع وصيف عام 1999 من قبل قسم علوم الكمبيوتر والإدارة بجامعة ترينتينو بإيطاليا. الدراسة هي جزء من مشروع أكبر هدفه الرئيسي هو تحديد مزايا وعيوب أبحاث السوق التي تتم عبر الإنترنت فيما يتعلق بالطرق والقنوات التقليدية ، والنظر في إمكانية تطبيقها في أسواق المنتجات المتنوعة. من الناحية المنهجية ، كان الهدف من البحث المقدم في هذه الورقة هو إظهار فوائد إجراء دراسات السوق عبر الإنترنت للمنتجات المبتكرة. تنبع المشاكل المتعلقة بهذه المنتجات المبتكرة أولا من حقيقة أنه لا يمكن تحديد خصائصها بدقة قبل إجراء البحث ، وثانيا ، يتطلب توافرها في شكل تجاري عادة المزيد من الاستثمارات الكبيرة في البحث والتجريب. كل من هذه القضايا حاسمة بالنسبة لأدوات CASE (هندسة البرمجيات بمساعدة الكمبيوتر) ، والتي تستخدم الأدوات اللغوية لتحليل الوثائق باللغة الطبيعية ، وبالتالي فهي تستند إلى تقنيات معالجة اللغة الطبيعية (NLP) التي تم تطويرها في مجال الذكاء الاصطناعي. من خلال العمل من منظور شركة تحاول تحديد المنتجات التي يجب تطويرها (من بين المشاريع المختلفة المتعلقة بالتطبيقات القائمة على البرمجة اللغوية العصبية) ، كان هدفنا هو تقييم الطلب المحتمل على أدوات CASE القائمة على البرمجة اللغوية العصبية. عند إجراء الدراسة ، وضعنا افتراضا معقولا بأن المستجيبين (الأشخاص المشاركين في تطوير أنظمة البرمجيات) يمكن الاتصال بهم بسهولة عن طريق الإنترنت. ولا يمكن ضمان هذا الشرط الأساسي أساسا على المستوى

الوطني بالنسبة للقطاعات الأخرى التي سبق دراستها (مثل السياحة أو التجارة الإلكترونية لمحلات البقالة). في الوقت نفسه ، كان من المتوقع وجود استعداد معين لعدم المشاركة في الدراسة ، سواء بسبب ضيق الوقت (لوحظ حتى في المقابلات الاستكشافية الأولية) أو بسبب مستوى التشبع المرتفع بالفعل. في الواقع ، تم تأكيد كل من هذه الافتراضات خلال البحث. ومع ذلك، فإننا نؤكد أن هذه الورقة تركز على نتائج المحتوى الفعلي للبحث، وفيما يلي نصف فقط الجوانب المنهجية ذات الصلة بتفسير النتائج التي تم الحصول عليها. 403938

## اهداف

كما ذكرنا سابقا ، كان الهدف من البحث هو تحليل الطلب المحتمل على أداة CASE التي تدمج الأدوات اللغوية كدعم لتحليل المتطلبات [2]. لإعطاء السياق الذي يمكن فيه تصميم مثل هذه الأداة واستخدامها ، تصف الفقرة التالية أولا دور اللغة الطبيعية في هندسة المتطلبات ثم تصنف التطبيقات الممكنة للأدوات اللغوية ، مع الإشارة إلى بنية نظام البرمجة اللغوية العصبية المثالي وإلى الأنشطة الأساسية الثلاثة لتحليل المتطلبات: الاستنباط ، النمذجة والتحقق من الصحة [3]. تشير أبحاث السوق لدينا بشكل أساسي إلى دعم النمذجة المفاهيمية ، وهو نشاط يتطلب الاستفادة من استخدام الأدوات اللغوية تصميم وحدة نمذجة. ويمكن دعم الأنشطة الأخرى بالوظائف الحالية لنظام البرمجة اللغوية العصبية، بمستويات متفاوتة من الأداء.

وقد تبين في وقت مبكر من الدراسة أن أيا من أدوات CASE التجارية لم تستغل الأدوات اللغوية لدعم نمذجة المتطلبات [4]؛ وهذا يعني ، بالتالي ، أن أبحاث السوق كانت تركز على منتج جديد لا يمكن تحديد ميزاته فيما يتعلق بالمنتجات الحالية المماثلة (تحليل المنافسة). ومع ذلك ، توجد العديد من المشاريع البحثية في هذا المجال ، وهي بمثابة شهادة على الاهتمام الكبير باستخدام الأدوات اللغوية في هندسة المتطلبات. الهدف المشترك هو إجراء تحليل لغوي لوثائق المتطلبات من أجل إنتاج نماذج مفاهيمية لها. من بين أحدث المشاريع ، على سبيل المثال ، يمكننا الاستشهاد بتلك الموصوفة في [8،9]. في حين أن المراجعة الكاملة خارج نطاق هذه الورقة ، تجدر الإشارة إلى كيفية تحليل المناهج المختلفة من خلال النظر في جانبين رئيسيين (اعتمادا على خصائص الأدوات اللغوية المعتمدة): 4241

- (c) مدى "طبيعية" لغة الإدخال ، والتي تخضع عادة لقيود تتعلق بالقواعد أو المفردات أو كليهما ؛
- (d) مقدار التدخل المطلوب من قبل المحلل من أجل معالجة النص "شبه الآلي" أو لتحديد العناصر الرئيسية للنمذجة المفاهيمية.

يركز المسح الموصوف في هذه الورقة على أول هذه النقاط ، وهي نقطة نعتبرها مهمة حيوية لأنه مهما كان النهج المعتمد ، فإن "طبيعية" اللغة تؤثر بشكل مباشر على مقدار الجهد اللازم لاستخراج معلومات مفيدة من الوثائق. أولا، كان من الضروري تحديد ما إذا كانت الوثائق التي تم جمعها في مرحلة استنباط المتطلبات بلغة طبيعية "حقيقية" أو بنوع من اللغة المقيدة، وإذا كانت بلغة طبيعية، ما إذا كان يمكن أن يطلب من المستخدم أو العميل وصف المتطلبات باستخدام لغة أكثر تقييدا. في الواقع ، إذا كانت المستندات مكتوبة بلغة "خاضعة للرقابة" (قيود على القواعد أو المفردات) ، فيمكن استخراج المعلومات باستخدام

<sup>38</sup> مشروع متعدد السنوات بتمويل من قسم علوم الكمبيوتر والإدارة بجامعة ترينتو.

<sup>39</sup> تم وصف بعض المقارنات المستمدة من بحثنا في [1].

<sup>40</sup> لمزيد من الدراسة للقضايا المتعلقة بأبحاث السوق عبر الإنترنت ، فإن المهمة يمكن أن تشير إلى الأدبيات (انظر على سبيل المثال ، المنشورات الموجودة في ESOMAR - الجمعية الأوروبية لأبحاث الرأي والتسويق - <http://www.esomar.nl>).

<sup>41</sup> انظر [5,6]. تتوفر ببلوغرافيا في <http://nl-oops.cs.unitn.it>.

<sup>42</sup> المقترحات الأولى لاستخدام المعايير اللغوية لاستخراج الكيانات والعلاقات ، ثم الأشياء والجمعيات ، من الأوصاف السردية للمتطلبات تعود إلى 1980s [7].

تقنيات نحوية أو "ضحلة" ، مثل تحليل الأشجار. للحصول على أداء مكافئ مع المستندات بلغة طبيعية غير مقيدة ، من الضروري أن يكون لديك ملف 43

التمثيل الدلالي للمعرفة التي تتضمن تقنيات التفكير. ويجري حاليا دراسة هذه التطبيقات. وعلاوة على ذلك، يمكن ربط اللغة المستخدمة في الوثائق بشكل أو بآخر بمجال تطبيق معين (على سبيل المثال، برمجيات الاتصالات السلوكية واللاسلكية)، وبالتالي تحديد درجة تخصص أداة الدعم اللغوية التي ستستخدم في التحليل المفاهيمي، وبالتالي في قاعدة معارفها. بمعنى آخر ، بافتراض أن تقنيات البرمجة اللغوية العصبية الأساسية متاحة ، بالنسبة للشركة التي يجب أن تقرر ما إذا كانت ستستثمر في تطوير أداة قائمة على البرمجة اللغوية العصبية لتحليل المتطلبات أم لا ، من المهم أن تحدد أولا ما إذا كان من الممكن تصميم وتحقيق أداة للأغراض العامة لدعم تطوير البرامج لمجالات التطبيقات المختلفة أو إذا كان من الضروري بدلا من ذلك إجراء المزيد من الاستثمارات لاحقا لتخصيص الأداة لمختلف الشركات أو العملاء الذين ستخدمهم في النهاية. هذه كلها اعتبارات أساسية في تحديد الاستثمار اللازم لتحويل نموذج بحثي - مثل تلك التي تم تطويرها في مشاريع البحث الحالية - إلى أداة تجارية. 44

قادتنا نتائج المقابلات الأولية بالإضافة إلى أحدث النماذج الأولية الحالية إلى اتخاذ قرار بعدم التحقيق في درجة تدخل المحلل المطلوب أو الأداء المطلوب للأداة (النقطة ب: نقتصر في هذه النقطة على إعطاء بعض النتائج العامة التي ظهرت أثناء إجراء البحث). وكان القيام بذلك يتطلب مزيدا من الاستثمار في إجراء بحوث سوقية أكثر شمولاً؛ ولن يكون من الممكن تبرير هذه الدراسة إلا بنتيجة إيجابية، ليست مضمونة بالتأكيد، فيما يتعلق بالمسائل المتصلة بالنقطة ألف). علاوة على ذلك ، لتقييم السوق المحتملة لأداة قائمة على البرمجة اللغوية العصبية لتحليل المتطلبات ، درسنا الجوانب المتعلقة بنشر أساليب وأدوات هندسة البرمجيات. على وجه الخصوص ، نعتزم التحقق مما إذا كان تحليل المتطلبات يعتبر في الواقع أمراً بالغ الأهمية فيما يتعلق بالأنشطة المهمة الأخرى في تطوير البرمجيات (الاختبار ، التوثيق ، إلخ).

## هيكل الورقة

تم تنظيم الورقة على النحو التالي: يصف القسم التالي سياق أداة CASE التي تدعم البرمجة اللغوية العصبية ويلخص التطبيقات الممكنة للأدوات اللغوية لهندسة المتطلبات. ويوفر ذلك معلومات عن تصميم الاستبيان والتفسير النهائي للنتائج. يحدد القسم الثالث خطة أبحاث السوق ، مع ملاحظة المراحل المختلفة والتركيز على الاستبيان وعلى خصائص المستجيبين. يتم عرض النتائج الرئيسية للمسح عبر الإنترنت في القسم الرابع ، حيث يتم تحليلها باستخدام تقنية إحصائية يشار إليها باسم تحليل المراسلات. كشفت الملفات الشخصية التي تم الحصول عليها عن وجود مكانتين في السوق تتميز بمناهجها المتنوعة لتطوير البرمجيات. أخيراً ، تم تقديم بعض الملاحظات فيما يتعلق بخصائص المسح وإمكانية توسيع النتائج. تلخص الاستنتاجات كيف يمكن استخدام نتائج المسح من قبل أولئك الذين يطورون البرامج بشكل عام ، وأولئك الذين يصممون الأدوات والبيئات لتحليل المتطلبات بشكل خاص.

## 2. دور اللغة الطبيعية في هندسة المتطلبات

لقد كتب الكثير عن أهمية تحليل المتطلبات. من أجل إظهار سبب كون البيئات والأدوات اللازمة لدعم هذا التحليل أقل إرضاء من تلك المتاحة للمراحل الأخرى من دورة حياة البرنامج ، سنستعرض بإيجاز السمات المميزة لهندسة المتطلبات ، والتي يتم تعريفها على النحو التالي:

<sup>43</sup> المدرجة في هذه الفئة ، على سبيل المثال ، instrumentTs الموصوفة في [10] و [11].

<sup>44</sup> على سبيل المثال، إلى الاعتراف لو واشنطن هو اسم شخص أو مطار أو مدينة في وثيقة معينة تتطلب نهجا دلاليا. ولا تسمح القيود المفروضة على الحيز بإجراء مناقشة أعمق لهذه المسألة هنا؛ حد ذاته على سبيل المثال [12].



"النهج المنهجي لتطوير المتطلبات من خلال عملية تعاونية تكرارية لتحليل المشكلة ، وتوثيق الملاحظات الناتجة في مجموعة متنوعة من أشكال التمثيل ، والتحقق من دقة الفهم المكتسب". [3 ، ص 13].

وبالتالي يتضح الأهمية المركزية للتواصل والمعرفة. بالمقارنة مع المراحل الأخرى لهندسة البرمجيات ، فإن تحليل المتطلبات والنمذجة المفاهيمية [15] يمثل صعوبات فريدة. العديد من الأنشطة المعنية معرفية وتتطلب الإبداع وكذلك المعرفة حول تكنولوجيا المعلومات ومجال التطبيق. وعلاوة على ذلك، فإن أوجه التقدم الأخيرة الناجمة عن إعادة هندسة العمليات التجارية وإدراج عناصر مبتكرة في نظم المعلومات تؤدي إلى توسيع نطاق المشاريع. ونتيجة لذلك، ازداد عدد الجهات الفاعلة والتفاعلات واللغات المعنية. وتكتمل الصورة باحتياجات الشركات، التي تعمل بمستويات أعلى من أي وقت مضى من القدرة التنافسية والتي تتطلب نظم معلومات مرنة بشكل متزايد.<sup>45</sup>

في هذا السياق، فإن استخدام الأدوات اللغوية – وبشكل أكثر دقة لأنظمة البرمجة اللغوية العصبية – لدعم تطوير أنظمة البرمجيات بشكل عام وتحليل المتطلبات بشكل خاص، قد يساعد المحلل على:

- التركيز على المشكلة بدلا من النمذجة؛
- التفاعل مع الجهات الفاعلة الأخرى ؛
- تأخذ في الاعتبار أنواع مختلفة من المتطلبات (التنظيمية والوظيفية ، وما إلى ذلك) ؛
- تحقيق إمكانية التتبع اعتبارا من الوثائق الأولى التي تم إنتاجها ؛
- إدارة مشكلة متطلبات المستخدم المتغيرة بشكل أكثر كفاءة.<sup>46</sup>

فيما يتعلق بالتطبيقات الممكنة لأنظمة البرمجة اللغوية العصبية لهندسة المتطلبات ، تجدر الإشارة إلى أنها قادرة على معالجة كل من المدخلات الصوتية والنصية ، وفي بعض الأحيان تفرض قيودا مثل الحد من المفردات أو القواعد.

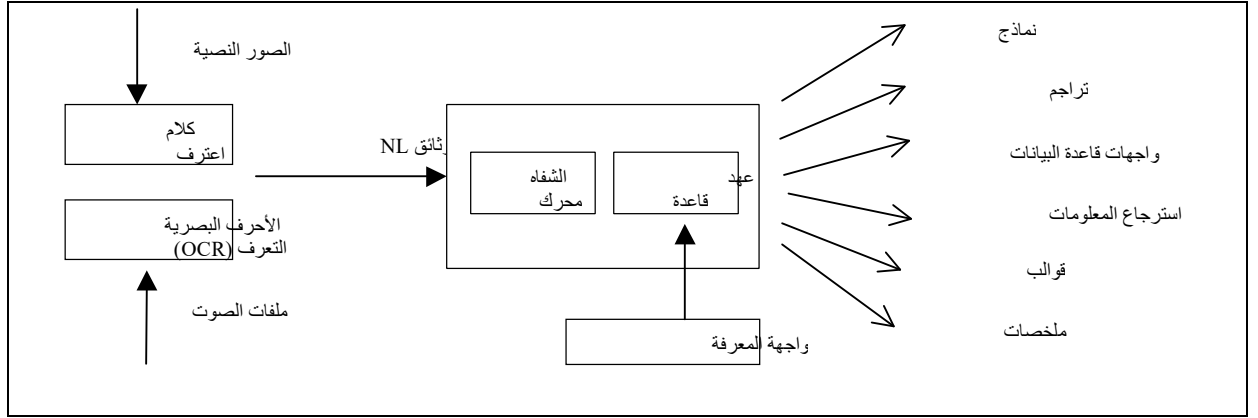
يمكن استخدام أنظمة البرمجة اللغوية العصبية للحصول ، بمستويات مختلفة من الأداء ، على ثلاثة أنواع من المخرجات:

- التحليل النحوي أو الدلالي أو العملي ؛
- النص إما بنفس اللغة أو لغة أخرى ، طبيعية أو اصطناعية ؛ - توليفات في شكل ملخصات أو قوالب منظمة بشكل مختلف.

الشكل 1 هو مخطط مبسط لنظام البرمجة اللغوية العصبية المثالي للأغراض العامة. من المهم أن نتذكر أن أنظمة التطبيقات الحقيقية عادة ما تعتمد بشكل كبير على المهمة وعلى المجال<sup>11</sup>.

<sup>45</sup> "الجزء الصعب ، والجوهر الحقيقي للمتطلبات ، هو محاولة فهم احتياجات عميلك. يحتاج الشخص المشارك في المتطلبات إلى مهارات بشرية ، ومهارات اتصال ، ومهارات فهم ، ومهارات شعور ، ومهارات استماع" [13]. انظر أيضا [14].

<sup>46</sup> للحصول على دراسة حديثة حول سبب استحالة معرفة المستخدمين لمتطلباتهم مسبقا ، انظر [16].<sup>11</sup> حول هذه النقطة ، انظر ، على سبيل المثال ، المهام التي تتطلبها مسابقات MUC (مسابقة فهم الرسالة) تنظم بواسطة الـ DARPA (وكالة مشاريع البحوث الدفاعية المتقدمة) [17].



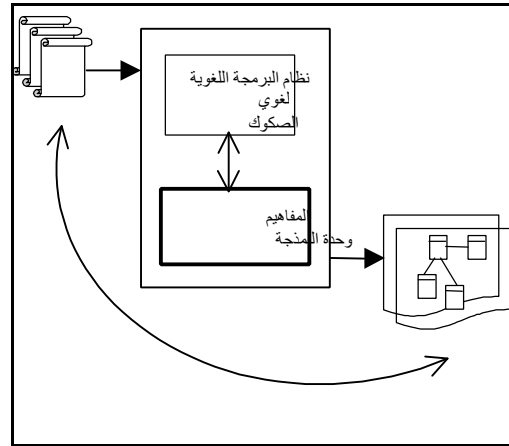
الشكل 1 - بنية نظام البرمجة اللغوية العصبية للأغراض العامة

بالإشارة إلى هذا المخطط ، يمكن استخدام الأدوات اللغوية ذات التعقيد المختلف وخاصة النضج المختلف:

(d) في مرحلة استنباط المتطلبات:

- تسهيل رقمنة وثائق المتطلبات باستخدام أنظمة التعرف على الكلام أو واجهات الاستجاب القائمة على البرمجة اللغوية العصبية ؛
- للكشف عن الغموض والتناقضات في الوثائق التي تصف احتياجات المستخدم (انظر على سبيل المثال ، [12،18،19] ؛
- لتصميم استبيانات أو مقابلات ، من خلال التحقق من غموض الأسئلة ؛
- للتحليل التلقائي للردود على الأسئلة المفتوحة ، وتفسير وتصنيف محتوياتها [20].

(e) لنمذجة المتطلبات عن طريق استخراج (مباشرة من النص) أوصاف العناصر لتضمينها في النماذج المفاهيمية المتوخاة في طريقة التطوير المعتمدة ، ولا سيما مخططات UML (لغة النمذجة الموحدة) (انظر الشكل 2).<sup>47</sup>



الشكل 2 - عملية توليد النماذج

<sup>47</sup> يمكن أن تكون الوثائق الرسمية لمواصفات UML وجد على موقع OMG (مجموعة إدارة الكائنات) على الويب:

<http://www.omg.org>

(f) لدعم التحقق من صحة المتطلبات ، من خلال استغلال وظيفة توليد أنظمة البرمجة اللغوية العصبية لإنتاج أوصاف باللغة الطبيعية بناء على الهياكل المستخدمة لتمثيل المعرفة.

تتطلب الرؤية الكاملة ملاحظة أنه يمكن أيضا استخدام أدوات البرمجة اللغوية العصبية للتوثيق، وإنشاء تقارير عن المراحل المختلفة لجمع المتطلبات ونمذجتها؛ ومن أجل إمكانية التتبع، مما يسمح بالحفاظ على رابط بين النصوص المستخدمة والنماذج المنتجة؛ ولترجمة الوثائق إلى لغات مختلفة، وهو أمر يصبح ضروريا بشكل متزايد في تصميم نظم المعلومات الدولية.

تتعلق الدراسة الاستقصائية الموصوفة في هذه الورقة بالنقطة الثانية من هذه النقاط ، أي استخدام تقنيات البرمجة اللغوية العصبية لدعم تطوير النماذج المفاهيمية ، نظرا لأنها تتطلب تصميم وحدة نمذجة. يمكن دعم جميع الأنشطة الأخرى من خلال الوظائف الحالية لنظام البرمجة اللغوية العصبية المثالي ، وإن كان ذلك بأداء مختلف. والافتراض الأكثر أهمية هو أن وثائق المتطلبات، بمجرد تحليلها، يمكن أن تسهم في "قاعدة معارف" تستخرج منها العناصر التي تعتبر مفيدة لأنشطة النمذجة. هناك جانبان مهمان يجب ملاحظتهما فيما يتعلق بمشاريع تطوير هذا النوع من الأدوات: (i) يعتمد العديد من هذه المشاريع على أنظمة/البرمجة اللغوية العصبية المخصصة ، وبالتالي لا يبدو أنها تتوافق مع متطلبات قابلية التوسع ومتانة التطبيقات الحقيقية ؛ (ب) نظرا لتعقيد اللغة الطبيعية ، يتوقع جميعهم تقريبا أن تكون الوثائق مكتوبة بلغة مقيدة أو أن بعض المراجعة للنص قد حدثت قبل الخضوع للتحليل التلقائي. هاتان الحقيقتان تستحقان التذكر عند تفسير نتائج أبحاث السوق وعند تقدير الاستثمارات المحتملة في تقنيات البرمجة اللغوية العصبية ، وبالتأكيد عند تطوير وحدة CASE لدعم تحليل المتطلبات.

### 3. تخطيط وتحقيق أبحاث السوق

تم اتخاذ قرار التحقيق في سوق أداة قائمة على البرمجة اللغوية العصبية لتحليل المتطلبات في سياق مشروع بحثي مشترك مع قسم علوم الكمبيوتر بجامعة دورهام (المملكة المتحدة) حيث تم تطوير نموذج أولي لأداة CASE - تسمى NLOOPS - لنمذجة المتطلبات وفقا للنهج الموجه للكائنات [21،22].<sup>48</sup>

استندت أبحاث السوق الموصوفة هنا إلى إدارة استبيان يتطلب تصميمه النظر في الخبرة المكتسبة طوال تطوير NLOOPS ، ومنهجية وتقنيات أبحاث السوق عبر الإنترنت. على وجه التحديد ، تقدم البحث في المراحل التالية:

- مسح أولي
- تحديد الأشخاص الذين أجريت معهم المقابلات
- تصميم واختبار الاستبيان
- اختيار طريقة الاتصال
- توزيع الاستبيان والتذكير - جمع وتحليل البيانات.

يتبع ذلك وصف لكل مرحلة ، مع التركيز بشكل أكبر على المرحلة الثالثة (تصميم الاستبيان) والمرحلة النهائية (تحليل البيانات).

كانت الخطوة الأولى في المشروع البحثي هي إنشاء مجموعة تركيز تتألف من كل من الشركات التي تطور الأدوات اللغوية وكذلك الشركات الكبيرة والصغيرة التي تطور البرمجيات أو تقدم خدمات مرتبطة بإدخال تقنيات المعلومات في مكان العمل. كان الهدف من هذه المرحلة هو جمع معلومات حول احتياجات المستخدمين التي يمكن تلبيتها باستخدام أداة CASE القائمة على البرمجة اللغوية العصبية وجمع معلومات أخرى مفيدة في تصميم الاستبيان. واجه الباحثون على الفور وجهات نظر متشائمة للأدوات التي تستخدم

<sup>48</sup> اللغة الطبيعية - نظام الإنتاج الموجه للكائنات ، <http://nl-oops.cs.unitn.it>.

تقنيات البرمجة اللغوية العصبية لدعم تحليل المتطلبات. وعلى وجه الخصوص، أعرب بعض أعضاء الفريق المتخصص عن شكوك جدية في أن اللغة الواردة في الوثائق التي تم جمعها لتحليل المتطلبات كانت "طبيعية" بما فيه الكفاية لتبرير اعتماد أداة تستند إلى تقنيات البرمجة اللغوية العصبية. وشكك آخرون في الجدوى التقنية لمثل هذه الأدوات، مشيرين إلى تجاربهم غير المرضية مع تطبيقات البرمجة اللغوية العصبية الأخرى مثل برامج الترجمة.

**تحديد الأشخاص الذين أجريت معهم المقابلات** ووفقا لهدف الدراسة، وجه الاستبيان أساسا إلى الأشخاص المشاركين في تطوير البرمجيات، بالإضافة إلى المديرين المسؤولين عن القرارات الهامة المتعلقة بعملية تطوير البرمجيات، بما في ذلك قرار اعتماد منهجيات وأدوات دعم. من وجهة نظر إحصائية، عند التعامل مع مسح يتم إجراؤه عبر الإنترنت، تتمثل إحدى المشكلات الرئيسية في تحديد الدرجة التي تمثل بها العينة السكان المستهدفين، وفي هذه الحالة الأشخاص أو الشركات المشاركة في تطوير البرمجيات. من ناحية، من المعقول افتراض أنه يمكن الوصول إلى المستجيبين المستهدفين عن طريق الإنترنت، بينما من ناحية أخرى يتمتع السكان بخصائص (العدد والحجم والتوزيع الجغرافي وما إلى ذلك) غير موثقة. بالنظر إلى هذا وبالنظر أيضا إلى طرق الاتصال المختارة، فإن نهج الدراسة يشبه من الناحية المفاهيمية أخذ العينات المتسلسلة. إحصائيا، هذا من شأنه أن يصنفها على أنها دراسة وصفية، وبالتالي يتطلب الحذر عند توسيع النتائج خارج عينة المسح.

**تصميم واختبار الاستبيان** وبالنظر مرة أخرى إلى أهداف الدراسة، من حيث المنهجية والمحتوى على حد سواء، لم تجر الدراسة الاستقصائية إلا عن طريق الإنترنت وتألقت من استبيان على صفحة على شبكة الإنترنت (انظر التذييل ألف). كان هذا الاختيار هو القوة الدافعة خلال مرحلة التصميم والاختبار، وكان الهدف هو الحصول على استبيان موجز مع أسئلة مغلقة بلغة واضحة قدر الإمكان. أما بالنسبة للأسئلة نفسها، فقد تم الاختيار مع ظهور قضايا منطقية وذات صلة طوال فترة عمل الفريق المتخصص. وبعد مرحلة من الاختبار خضع فيها الاستبيان للتدقيق - مباشرة أولا ثم عبر الإنترنت - لمجموعة مختارة من المحللين ومديري المشاريع، تم إنتاج النسخة النهائية. تم تقسيم الاستبيان النهائي إلى قسمين، لما مجموعه ثمانية عشر سؤالا، وسؤال نهائي مفتوح لمزيد من الملاحظات. تألفت المجموعة الأولى من أسئلة تتعلق بالشركة (الأسئلة 1-4) وبالمدعى عليه (السؤالان 5 و 6). أما الجزء الثاني فقد بحث في عمليات إنتاج البرمجيات، بحيث تتعلق مجموعة واحدة من الأسئلة باستخدام المنهجيات (الأسئلة 7-10) والأدوات (السؤالان 13 و 14) في تطوير البرمجيات؛ وتناولت مجموعة أخرى الوثائق المستخدمة في تحليل المتطلبات (الأسئلة 11 و 12 و 15) وكانت الثلاثة الأخيرة تتعلق بكفاءة عملية التطوير (الأسئلة 16 و 17 و 18). سئل المستجيبون أيضا عما إذا كانوا مهتمين بالحصول على نتائج البحث أو في مشاهدة عرض توضيحي لنموذج أولي لأداة CASE القائمة على البرمجة اللغوية العصبية. تم اتخاذ قرار إدخال الأسئلة المرتبطة بنهج هندسي لتطوير البرمجيات بعد التحقق من إمكانية استخدام البيانات الموجودة. والمثير للدهشة أنه لم يتم العثور إلا على كمية صغيرة من البيانات، سواء لنشر المنهجية الموجهة للكائنات أو لاستخدام النماذج "الكلاسيكية" مثل نماذج علاقات الكيانات. هذه مهمة لأن البحث المبكر والنماذج المفاهيمية للتحليل اللغوي للمتطلبات [7] بحثت في إنتاج مخططات علاقات الكيان. علاوة على ذلك، يمكن اعتبار هذه النماذج حالة خاصة من النماذج الطباقية التي يتوقعها النهج الموجه للكائنات. فيما يتعلق بسوق أدوات CASE، في كثير من الحالات لم تلبى التوقعات ونتيجة لذلك لم تحقق النجاح المطلوب في السوق [25]. سيتعين علينا انتظار اعتماد UML - الذي تم تطويره قبل حوالي عام واحد من بدء المشروع البحثي الحالي - كمعيار للنمذجة المفاهيمية من قبل OMG (مجموعة إدارة الكائنات)؛ عندها فقط سيكون هناك نمو كبير في سوق أدوات CASE، التي أعيد تجميعها وإعادة تسميتها كأدوات نمذجة الكائنات أو أدوات النمذجة المرئية. باختصار، أثرت ندرة البيانات حول تغلغل ودور النهج الهندسي لتطوير البرمجيات

على اختيار الأسئلة للمسح ، ولكن أيضا ، كما سنرى ، القدرة على التحقق من صحة النتائج وتوسيعها.  
52515049

الأسئلة التي تعتبر الأكثر أهمية للتحقق من وجود مكانة سوقية لأداة CASE القائمة على البرمجة اللغوية العصبية هي تلك المتعلقة بالمستندات المستخدمة لجمع المتطلبات. في الواقع ، كما رأينا بالفعل ، إذا كانت المستندات في اللغة الإنجليزية الحقيقية ، فهناك حاجة إلى تقنية أكثر تطورا (وتكلفة) لتطوير بيئة تدعم التحليل بشكل فعال باستخدام الأدوات اللغوية. لذلك من المفيد تحديد ما إذا كانت الشركة في وضع يمكنها من مطالبة العملاء أو المحللين بوصف المتطلبات بلغة مفيدة. يمكن أن تتعلق القيود النموذجية بما يلي: أ) القواعد - تهدف إلى الحصول على إنشاءات نحوية يسهل تحليلها من خلال طلب ، على سبيل المثال ، عبارات أقصر ، باستخدام الصوت النشط ، عن طريق تجنب المراجع المجازية ، وما إلى ذلك ؛ ب) المفردات - تهدف إلى تقليل غموض المصطلحات. علاوة على ذلك ، من أجل تحديد درجة التخصيص المطلوبة لأداة محتملة قائمة على البرمجة اللغوية العصبية ، تناولت أسئلة أخرى مستوى تخصص المصطلحات ومعرفة المجال المطلوبة لتطوير البرنامج.

في الأسئلة المتعلقة بكفاءة عمليات الإنتاج ، سئل المستجيبون على وجه الخصوص عن التحسينات التي يرغبون في رؤيتها (الاختيار من قائمة من ثمانية أنشطة محتملة تعتبر حاسمة ، اثنان منها أساسيان لمرحلة تحليل المتطلبات) وكيف يمكن تحقيقها ، والاختيار من بين "النفيوض الداخلي" ، "الاستعانة بمصادر خارجية" و "الأتمتة". تم تصميم السؤال الأخير للتأكد مما إذا كانت الشركة قادرة على تسليم أنظمة البرمجيات أو المنتجات دون تأخير. وأخيرا ، وتمشيا مع القاعدة العامة لبحوث السوق ، تم تقديم حافز للمشاركة في شكل سحب عشوائي بين المستجيبين للحصول على تذاكر لأداء الأوبرا في أرينا في فيرونا.<sup>53</sup>

**اختيار طريقة الاتصال تتطلب أهداف البحث وخصائص الأداة بطبيعتها طريقة اتصال تسمح باستخدام الفعال للوقت والموارد بينما تصل في نفس الوقت إلى أكبر عدد من المستجيبين المحتملين.** في هذه النقطة ، لمراعاة حقيقة أن هناك مستوى عال من التشعب - بسبب العدد الكبير من طلبات الاستطلاع التي يتلقاها المستجيبون - فكرنا في البداية في إرسال الاستبيان إلى بعض مجموعات الأخبار المتخصصة ، مما يسلط الضوء على الطبيعة الأكاديمية للبحث. في المرحلة الأولى حددنا ثلاث مجموعات إخبارية يرتبط عملها بموضوع البحث (comp.object ، comp.software-eng ، alt.comp.software-tools) ؛ تمت إضافة عشرين مجموعة إخبارية أخرى لاحقا إلى القائمة (القائمة الكاملة متاحة في <http://online.cs.unitn.it>). ومع ذلك ، بعد أن أثبتت طريقة الاتصال هذه أنها أقل نجاحا مما كان متوقعا ، قررنا الاتصال بالشركات مباشرة عن طريق البريد الإلكتروني ، وتزويدهم بعنوان صفحة الويب حيث يمكنهم العثور على الاستبيان وإكماله. تم الحصول على عناوين الشركات عبر الإنترنت باستخدام محركات البحث ، ولا سيما دليل Yahoo!<sup>215554</sup> (مطورو برامج < < الكمبيوتر).

<sup>49</sup> الاستبيان متاح جنبا إلى جنب مع البيانات التي تم جمعها وغيرها من المواد البحثية ذات الصلة في <http://online.cs.unitn.it>.

<sup>50</sup> على سبيل المثال ، يجب تغيير استبيان مثل الاستبيان المستخدم في المسح الموصوف في [23] بشكل جذري ليتم استخدامه على الخط.

<sup>51</sup> في ضوء الملاحظات الواردة في [24] ، قد لا يكون هذا مفاجئا.

<sup>52</sup> تم اختيار أدوات السؤال 14 على أساس بيانات المبيعات لفترة سابقة للدراسة.

<sup>53</sup> لأن المسح انتهى في نهاية الساحة أوبرا useaso ، تم استبدال التذاكر بأقراص مدمجة لموسيقى الأوبرا بواسطة Verdi.

<sup>54</sup> في الواقع ، كان أحد أهداف الاستطلاع هو التحقق في الظروف التي يمكن بموجبها استخدام مجموعات الأخبار لإجراء استطلاعات عبر الإنترنت.

<sup>55</sup> عدد محدود من الاستبيانات التي تم الحصول عليها (44) و اتهامات بالبريد العشوائي.<sup>21</sup>

<http://www.yahoo.com>

**توزيع الاستبيان والتذكيرات** كما هو موضح أعلاه ، تم إدارة الاستبيان بطريقتين مختلفتين. في المرحلة الأولى ، تم الإعلان عنها على عدد من مجموعات الأخبار المخصصة لتطوير البرمجيات (مما أدى إلى إكمال 44 استبياناً و 39 شركة برمجيات) وفي المرحلة الثانية ، تم إرسال طلبات المشاركة في الاستطلاع عبر البريد الإلكتروني إلى 1541 عنواناً يتوافق مع 1234 شركة برمجيات. وبهذه الطريقة الثانية، تم الحصول على 107 استبيانات مكتملة تقابل 103 شركات. للحصول على هذه النتائج ، كان من الضروري في كثير من الحالات إرسال رسالة لتذكير المتلقي بالمشاركة في الدراسة ، ولكن في نفس الوقت السماح له أو لها بشرح قرار عدم إكمال الاستبيان. وكثيراً ما أشارت الأسباب المقدمة لعدم ملء الاستبيان إلى ضيق الوقت والعدد الكبير من الطلبات من هذا النوع الواردة (يمكن الوصول إلى رسائل البريد الإلكتروني المرسلة على الإنترنت على <http://on-line.cs.unitn.it>). بالإضافة إلى ذلك ، كانت العديد من العناوين غير صحيحة ، على الرغم من أن النسبة كانت منخفضة إلى حد ما (7.6٪ ، 6.1٪ إذا تم حسابها حسب عدد الشركات). وبالتالي ، بلغ عدد جهات الاتصال الصحيحة 1424 ، أي ما يعادل 1159 شركة.

56

**جمع البيانات وتحليلها** تم إرجاع ما مجموعه 151 استبياناً ، 91٪ في غضون خمسة أيام من إرسال الطلب الأولي أو الاستبيان نفسه. كان معدل الاستجابة المحسوب للاستبيانات المرسلة عبر البريد الإلكتروني حوالي 8٪. يمكن اعتبار ذلك نتيجة مرضية عند مقارنته بالدراسات الاستقصائية التقليدية التي يتم إجراؤها عن طريق البريد أو الفاكس ، ومع الدراسات الاستقصائية الأخرى لتطوير البرمجيات ، والتي كان معدل الاستجابة لها 3٪ [25]. من الناحية الإحصائية البحتة ، لا يمكن اعتبار مجموعة الشركات التي تم الاتصال بها - بينما تشكل في حد ذاتها عدداً كبيراً - كعينة تمثيلية من سكان شركات تطوير البرمجيات. وبالنظر إلى ذلك، من المهم تفسير النتائج بطريقة وصفية، مما يتطلب توخي الحذر في توسيعها. ومع ذلك ، سنرى أنه بالنسبة لبعض الأسئلة ، يمكن تقييم جودة نتائج الاستطلاع من خلال مقارنتها بتلك التي تم الحصول عليها من استطلاعات أخرى ومع البيانات المتعلقة بسوق CASE. وترد نتائج هذه المقارنات في نهاية الفقرة التالية. 57

وعلى المستوى المنهجي، أكد استخدام مجموعات الأخبار أن هناك القليل من الجهد المطلوب لمطالبة المستجيبين بالمشاركة، ولكن العدد المنخفض من الاستبيانات المكتملة قد يبطل هذه الميزة. علاوة على ذلك ، يجب تقييم استخدام مجموعات الأخبار على أساس العوامل التالية: مستوى التخصص وعدد الرسائل ووجود المشرف. في ضوء نتائج استطلاعنا ، في حالة مجموعات الأخبار المتخصصة للغاية ، حتى لو كانت محتويات الاستطلاع ذات صلة بها ، من أجل زيادة معدل الاستجابة ، ينصح بطلب موافقة المشرف ، أو تحديد واحد أو أكثر من قادة مجموعات الأخبار الذين يمكنهم إضفاء الشرعية على الاستطلاع بمشاركتهم. 58

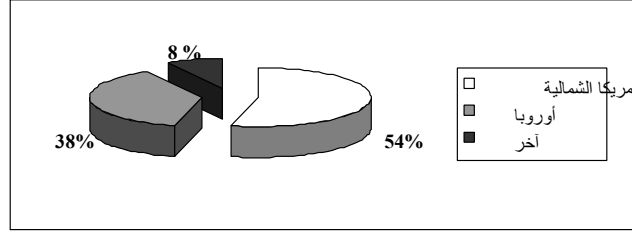
لاحظ التحليل الأولي التوزيع الجغرافي للمستجيبين، ومعظمهم من المقيمين في الدول الأوروبية أو أمريكا الشمالية (انظر الشكل 3). هذه النتيجة الأولى للبحث مدعومة بتحليل أوجه التشابه بين التوزيعات الجغرافية المختلفة (باستخدام المؤشرات المناسبة) مما يدل ، في الواقع ، على أن هذه الأسواق لها خصائص متشابهة.

56 هذه نسبة عالية إلى حد ما ، مع الأخذ في الاعتبار أنه تم جمعها من الصفحات الرئيسية لمواقع الشركة الرسمية. مسح آخر أجري في نفس الفترة على السياحة الشتوية ، حيث الأفعتم توفير ESSES من قبل مجلة متخصصة ، وجدت نسبة مماثلة جداً من العناوين الخاطئة (8.9٪) ، ولكن المبلغ يمكن أن يكون أعلى من ذلك بكثير. على سبيل المثال ، في دراسة استقصائية لمستخدمي الإنترنت أجريت في عام 1996 ، تم العثور على 35٪ من إجمالي 1221 عنواناً على أنهاز [26].

57 كان هذا هو الحد الأدنى لقيمة الدراسات الاستقصائية من النوع التقليدي ، والتي حققت معدل استجابة أقصى قدره 20٪. في الاستطلاع الذي وصفه جلاس وهوارد [25] ، ارتفعت النسبة المئوية إلى 17٪ بعد استكمال رسائل الاستبيان البريدية بواسطة teleph واحد الاتصالات مع متابعة الفاكس.

58 بالنسبة لمسح حول محلات السوبر ماركت الافتراضية ، تم إرسال رسالة إلى 6 مجموعات إخبارية للحصول على 100 استبيان مكتمل.

وبالنظر إلى ذلك، نقدم هنا نتائج الدراسة الاستقصائية بالكامل، مع تسليط الضوء فقط على الجوانب التي أثرت فيها منطقة الإقامة الجغرافية على الردود.



الشكل 3 - المجبيون حسب منطقة الإقامة الجغرافية

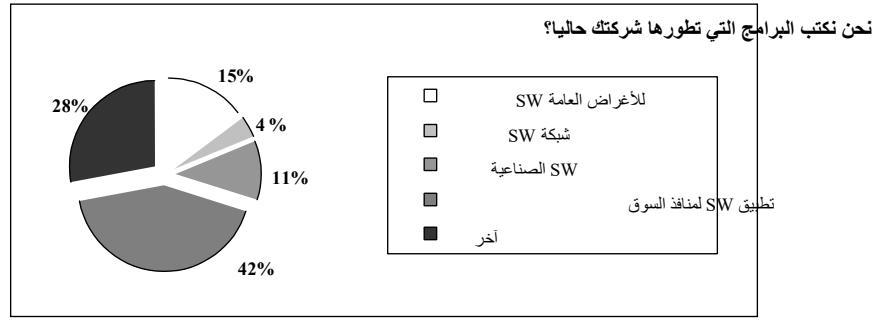
ويشغل ستة وثمانون في المائة من المشاركين أدوارا تتعلق بمشاريع تطوير البرمجيات، وشغل 68% منهم هذا الدور لأكثر من ست سنوات. وعلاوة على ذلك، وكما هو متوقع، فإن طول مدة الخدمة يؤثر على الوظيفة التي يشغلها في الشركة، بحيث يتم تنفيذ أعمال البرمجة بشكل أكثر تواترا من قبل الأشخاص العاملين لأقصر الفترات، في حين أن أولئك الذين عملوا في شركاتهم لمدة 6-10 سنوات تم توزيعهم بشكل موحد تقريبا بين الأدوار. تجدر الإشارة إلى أن غالبية المستجيبين الأوروبيين اختاروا "مهندس نظام / مهندس معماري" لكن نظرائهم الأمريكيين اختاروا "مدير المشروع"، والذي قد يكون بسبب استخدام مصطلحات مختلفة للدلالة على نفس الدور في المجالين. وعمل نحو 29% من المجبيين في شركات تضم أكثر من مائة موظف، على الرغم من أن الشركات الصغيرة كانت ممثلة تمثيلا جيدا (الجدول 1).<sup>59</sup>

لدي ما هو وجود العديد من الموظفين والاستشاريين في شركتك؟				
أكثر من 100	100-51	50-21	20-6	5-1
29%	5%	15%	24%	27%

الجدول 1 - حجم الشركة

العمل الأساسي للشركات التي شملها الاستطلاع في 77% من الحالات هو "تطوير SW" وفي 23% هو "مواقع الويب" أو "أخرى". كما هو متوقع، فإن أعلى نسبة من الشركات العاملة في أنواع أخرى من الأعمال (أو بالأحرى، أيضا في أنواع أخرى من الأعمال) تتألف من شركات أكبر حجما. وفيما يتعلق بنوع البرمجيات المنتجة، طورت 42 في المائة من الشركات برمجيات للأسواق المتخصصة (الشكل 4)، مع نسبة عالية تبلغ 48 في المائة لأمريكا الشمالية. قد يكون هذا بسبب وجود عدد أكبر من الشركات الصغيرة الحجم، بالنظر إلى أن 59% من الشركات التي لديها خمسة موظفين أو أقل، و 24% من تلك التي لديها أكثر من 100، تعمل في أسواق متخصصة. تم بيع منتجات البرمجيات في الغالب للمستخدم النهائي: 84%؛<sup>26</sup> تم بيع 13% فقط لشركة برمجيات أخرى، و 3% لمحلات البرمجيات. ومن المثير للاهتمام أن جميع الشركات التي طورت مواقع الويب باعت منتجاتها مباشرة إلى المستخدمين النهائيين، نظرا لطبيعة هذا النوع من المنتجات.

<sup>59</sup> تم حساب جميع النسب المئوية على إجمالي عدد المستجيبين الذين أجابوا على السؤال النسبي NS، مع حذف عدم الردود. <sup>26</sup> وسيطلب إجراء مزيد من التحقيق في هذا الجانب معرفة عدد وحجم عملاء الشركات. هذا، مع ذلك، خارج نطاق دراستنا الاستقصائية.



الشكل 4 - نوع البرنامج

تقدم الفقرة التالية تحليلاً مفصلاً لنتائج البحث في وجود سوق محتمل لأداة مبتكرة لدعم التحليل المفاهيمي - وهي أداة لديها القدرة على تحليل المستندات المكتوبة بمستويات مختلفة من اللغة الطبيعية.

4. نتائج المسح والطلب المحتمل على أداة قائمة على البرمجة اللغوية العصبية لدعم تحليل المتطلبات يمكننا تحديد ثلاث مجموعات من العناصر المفيدة في تقييم الطلب المحتمل على أداة CASE لدعم تحليل المتطلبات للمستندات المكتوبة باللغة الطبيعية. يمكن وصفها على النحو التالي ، مع مراعاة ترابطها: <sup>60</sup>

[4] سوق الأدوات التي تدعم تطوير البرمجيات ونمذجة المتطلبات. ما مدى اتساع السوق؟ ما مقدار المنافسة هناك؟ هل يستخدم مطورو البرامج أدوات CASE؟ إذا كان الأمر كذلك ، أي منها؟ (عادة ما يفترض استخدام أداة CASE اعتماد منهجية تطوير). كانت هذه النقطة الأخيرة مهمة لتحديد النماذج المفاهيمية التي يجب أن تدعمها الأداة (وهو جانب أصبح أقل أهمية مع نشر UML) ، ولأسباب تتعلق بالتوافق والتكامل مع الأدوات الحالية. ويمكن الحصول على بعض المعلومات عن هذه النقطة عن طريق البيانات المتعلقة بمبيعات أدوات CASE ، ولكن أدرج سؤال واحد بشأن هذا الموضوع يتعلق بالأدوات الداعمة لتحليل المتطلبات وتصميم المستوى الأعلى. <sup>6261</sup>

[5] مميزات الأداة. المتطلبات التي تؤثر بشكل أساسي على الاستثمارات اللازمة لتطوير أداة لتحليل المتطلبات على أساس الأدوات اللغوية هي (أ) اللغة الموجودة في الوثائق التي تم جمعها في مرحلة استنباط المتطلبات ، والتي تعتبر حاسمة في تحديد التقنيات والأدوات اللغوية المناسبة ، و (ب) درجة المعرفة المتخصصة بالمجال المطلوب من الأداة ، والتي تحدد درجة التخصص المطلوبة من منتج أداة CASE (العمومية). أيضاً ، بالنظر إلى أحدث الأدوات اللغوية ، فإن أحد الاعتبارات المهمة هو الأداء المطلوب من الأداة ؛ وبعبارة أخرى ، ما مدى "جودتها" لتستحق الشراء؟ <sup>63</sup>

[6] تحليل المتطلبات ينظر إليه على أنه حاسم. وهذا عنصر حيوي في تحديد المنافذ السوقية المحتملة وفي التأكد من إمكانية قيام المستخدمين بالاستثمار في أداة تدعم تحليل المتطلبات، فضلاً عن استعدادهم وقدرتهم على قبول التغييرات التي تصاحب اعتماد أداة جديدة. الشركات التي لديها نهج هندسي لتطوير البرمجيات لديها عمليات موحدة للغاية ، وبالتالي يجب أن تعتبر الأنشطة التي تفنر إلى الهيكل أو الدعم

<sup>60</sup> للحصول على مقدمة لتقييم deman المحتملة ، انظر على سبيل المثال ، [27].

<sup>61</sup> في الماضي ، كانت الحاجة إلى دعم الرموز الرسومية المختلفة عيباً في سوق CASE ، من حيث أنها تتطلب من المنتجين اختيار الترميز الذي يجب دعمه بأدواتهم الخاصة ، أو استيعاب التكلفة الأعلى للتطوير في إصدارات مختلفة.

<sup>62</sup> لا يتطلب CASE القائم على التقنيات اللغوية للتحليل الموجه للكائنات بالضرورة تحقيق من بيئة دعم كاملة ، ولكن يمكن اعتبارها وحدة نمطية يمكن دمجها مع منتج موجود.

<sup>63</sup> تعتبر دراسة "المناعة" ذات أهمية قصوى أيضاً لتحديد درجة تدخل المحلل المطلوبة في تطوير نماذج المتطلبات ، ويجب إجراؤها باستخدام نموذج أولي للأداة. انظر أيضاً النقطة (ب) من المقدمة والخاتمة. الاستخدام.



نقاطا حاسمة تتطلب الاهتمام. إن الشركة التي تستخدم عملية غير رسمية أو "حرفية" لن تشارك بالضرورة هذا القلق ولكنها ستكون أكثر اهتماما باستخدام اللغة الطبيعية.

ولجمع المعلومات الأكثر فائدة بشأن هذه النقاط الثلاث، قمنا بتحليل الاستبيانات المكتملة على مرحلتين. في المرحلة الأولى نظرنا في الإجابات الفردية ، ودراسة العلاقات المتبادلة والتبعيات. في المرحلة الثانية ، طبقنا تحليل المراسلات [28] ، بهدف الكشف عن وجود ملفات تعريف تتوافق مع منافذ السوق المحتملة لأداة CASE المبتكرة.

[1] أما بالنسبة لاستخدام أداة تدعم تحليل المتطلبات وتصميم المستوى الأعلى ، فقد أجاب 30٪ فقط بشكل إيجابي. كما كان متوقعا ، تم استخدام هذه الأدوات بشكل أكبر في الشركات كبيرة الحجم ، حيث وصلت إلى 51٪ في تلك التي لديها أكثر من مائة موظف ، كما هو موضح في جدول التوزيعات المشروطة (الجدول 2). ليس من المستغرب أن يزداد استخدام هذه الأدوات مع طول مدة الخدمة (ترتفع من 17٪ إلى 36٪) مع المحللين كفئة من الموظفين الذين يستخدمونها بشكل متكرر.

الجدول 2 - استخدام أدوات تحليل المتطلبات وتصميم المستوى الأعلى حسب حجم الشركة

كم عدد الموظفين والاستشاريين في شركتك؟					
هل تستخدم أي أداة تدعم تحليل المتطلبات وتصميم المستوى الأعلى؟	5 - 1	6 - 20	21 - 50	51 - 100	أكثر من 100
نعم	16%	18%	33%	33%	51%
لا	84%	82%	67%	67%	49%

علاوة على ذلك ، ذكر 84٪ من المستجيبين أنهم استخدموا منهجيات محددة لتطوير البرمجيات. كان الحجم سمة مميزة هنا ، 78٪ من الشركات التي لديها خمسة موظفين أو أقل تستخدم منهجيات محددة و 93٪ لتلك التي لديها أكثر من 100. لا يؤثر نوع البرنامج أو قناة المبيعات بشكل كبير على استخدام المنهجيات ، على الرغم من أن الدور والخبرة يبدو أنهما يعلان ذلك إلى حد ما.

تم استخدام أفضل الرسوم البيانية المعروفة لنمذجة البيانات ، مخططات العلاقة بين الكيانات (E-R) من قبل 63٪ من المستجيبين الذين تبنوا منهجية. علاوة على ذلك ، يتوافق حجم الشركة الأصغر مع استخدامها الأكثر ندرة (52٪ في الشركات التي لديها أقل من خمسة موظفين ، و 73٪ في تلك التي لديها أكثر من 100). كان استخدام مخططات E-R أكبر بكثير بين المستجيبين الذين عملوا لفترة أطول في مجال الكمبيوتر (ارتفع من 35٪ بين أولئك الذين عملوا في هذا المجال لمدة تقل عن ثلاث سنوات إلى 66٪ بين أولئك الذين عملوا في هذا المجال لأكثر من عشر سنوات). وأخيرا، فيما يتعلق بنوع البرمجيات، استخدمت مخططات E-R بدرجات مختلفة جدا من قبل المجيبين الذين طوروا برمجيات للأغراض العامة (93٪) وأولئك الذين طوروا برامج الشبكات (25٪)، في حين لم تكن هناك اختلافات جوهرية فيما يتعلق بالبنود الأخرى.

كانت النسبة المئوية للمستجيبين الذين استخدموا طريقة موجهة للكائنات (OO) 68٪ ، وهي نسبة مماثلة لمستخدمي مخطط E-R. يظهر التصنيف حسب حجم الشركة اختلافا بين الشركات التي لديها خمسة موظفين أو أقل (60٪ منها تستخدم أساليب OO) وتلك التي لديها أكثر من 100 (74٪ منها تفعل ذلك). لا توجد اختلافات كبيرة فيما يتعلق بسنوات الخبرة ، في حين أن هناك ارتباطا أوثق بالوظيفة التي تشغلها الشركة: تراوحت النسب المئوية من 45٪ للمبرمجين إلى 78٪ لمهندسي الأنظمة / المهندسين المعماريين. يمكن إجراء مقارنة مثيرة للاهتمام في الجدول 3 ، حيث يلاحظ المرء أن أولئك الذين يتبنون أساليب OO

كانوا معتادين بالفعل على استخدام مخططات E-R ، مما يشير إلى أنهم بدوا أكثر ميلا لاستخدام نهج OO.

الجدول 3 - مخططات العلاقة بين الكيان والطرق الموجهة للكائنات

هل تستخدم الرسوم التخطيطية لعلاقة الكيان لنمذجة متطلبات البيانات الخاصة بك؟		هل تستخدم طريقة OO؟
لا	نعم	نعم
63%	69%	نعم
37%	31%	لا

فيما يتعلق بطريقة OO الأكثر استخداما ، أعلن 77٪ من المستجيبين الذين أجابوا بالإيجاب على السؤال السابق أنهم يستخدمون UML. هذه نتيجة تؤكد تأكيد UML كمعيار صناعي لنمذجة OO. تجدر الإشارة إلى أن المسح تم إجراؤه بعد عام ونصف تقريبا من اعتماد UML من قبل OMG.

وتبين أيضا أن الغالبية العظمى من المجيبين الذين قالوا إنهم لم يستخدموا المنهجيات لم يستخدموا أدوات لتحليل المتطلبات وتصميم المستوى الأعلى أيضا (90٪): في الواقع ، هناك ارتباط بين استخدام المنهجيات وأدوات CASE. ومن النتائج الأخرى التي ينبغي التأكيد عليها العلاقة بين استخدام أدوات CASE لتحليل المتطلبات أو تصميم المستوى الأعلى ونوع اللغة المستخدمة في الوثائق. ليس من غير المتوقع أن يتم استخدام هذه الأدوات بشكل متكرر عندما كانت اللغة أكثر رسمية (24٪ مع "لغة طبيعية مشتركة" و 63٪ مع "لغة رسمية"). حتى لو كان ينبغي التعامل مع هذه النتائج بحذر ، نظرا لانخفاض عدد الشركات التي شملها الاستطلاع ، فإنها تؤكد على ما يبدو عدم قدرة أدوات CASE المتاحة حاليا على تلبية احتياجات معالجة اللغة الطبيعية من خلال توفير بيانات مفيدة بشكل فعال. فيما يتعلق بالأدوات المستخدمة ، صرح 52٪ من المستجيبين الذين أجابوا بالإيجاب على السؤال السابق أنهم استخدموا Rational Rose. كانت Rational Rose الأداة ذات أعلى حصة في السوق في جميع أنحاء العالم وفي أوروبا. في عام 1998 ، كانت تمثل 33٪ من السوق ، بزيادة قدرها 79٪ عن العام السابق. ولهذا السبب ، يبدو أن النسبة المئوية التي وجدناها في الدراسة الاستقصائية التي أجريناها (52 في المائة لعام 1999) هي كما يتوقع المرء. 67666564

[2] كما لوحظ ، يحدد نوع اللغة المستخدمة في وثائق المتطلبات مدى تعقيد الأدوات اللغوية وتقنيات البرمجة اللغوية العصبية التي سيتم استخدامها. عندما تتم كتابة المستندات بلغة مقيدة (مجموعة فرعية من NL) - والتي تفرض قيودا على القواعد أو المفردات ، أو كليهما - يمكن استخدام أدوات لغوية أبسط وأكثر نضجا. ومع ذلك ، ليس من الممكن عادة فرض قيود على اللغة المستخدمة. أولا ، لأنه من الضروري اعتماد نهج

<sup>64</sup> لم يتم تحديد أي من الأدوات التي أشار إليها أولئك الذين يختارون الخيار "أخرى" أكثر من مرتين.

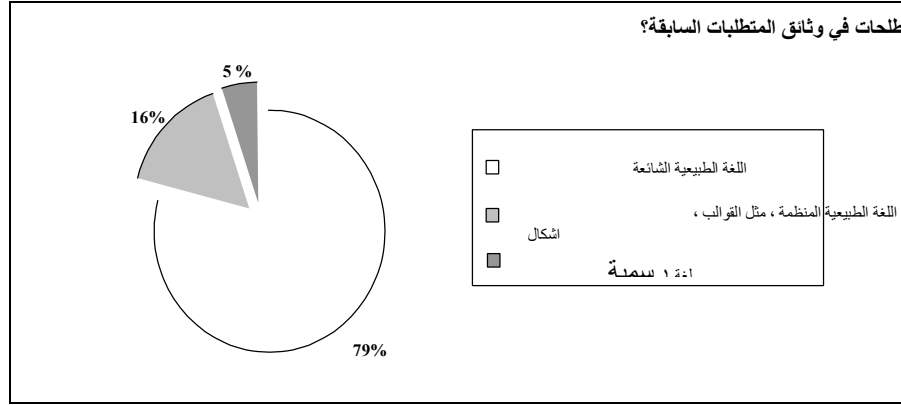
<sup>65</sup> بيانات مؤسسة البيانات الدولية (IDC).

<sup>66</sup> يبدو أن هذه الأرقام تتناقض مع نتائج الاستطلاع الذي أجراه Glass and Howard [25] ، حيث توصف تقنيات CASE بأنها في انخفاض. ومع ذلك ، تجدر الإشارة إلى أنه عندما يتعلق الأمر بـ CASE الخلفية أو "السفلية" ، فإن العديد من الوظائف معطلة الأحمر من قبل هذه الأدوات هي الآن جزء من بيئة التنمية. علاوة على ذلك ، غالبا ما يتم استخدام تعبيرات أخرى بدلا من "CASE": على سبيل المثال ، تستخدم استطلاعات IDC أدوات OOAMDC (التحليل الموجه للكائنات والنمذجة والتصميم والبناء). من جهة أخرى ، في عام 1998 نما سوق OOAMDC بأكثر من 10٪ (24٪ في أوروبا) ، انظر أيضا النتائج في [29].

<sup>67</sup> ومع ذلك ، تجدر الإشارة إلى أن بيانات الدراسة الاستقصائية التي أجريناها يتم التعبير عنها من حيث وحدات الإنتاج من قبل الشركات التي شملتها الدراسة الاستقصائية ، في حين أن أرقام المبيعات تحسب على أساس الفواتير وبالتالي تعتمد على الأسعار التي يتقاضاها البائعون.

موجه نحو العملاء في تطوير تطبيقات البرمجيات. ثانيا ، لأنه من الضروري تقليل خطر أن القيود المفروضة على اللغة والشكليات المعتمدة ستجبر المستخدم ، أو حتى المحلل ، على التعبير عما تسمح النماذج بتمثيله ، بدلا من المتطلبات الحقيقية للنظام. ويظهر الاستطلاع أنه في كل من أوروبا وأمريكا الشمالية، يتم تقديم وثائق المتطلبات مباشرة من قبل العميل وتتكامل مع المقابلات في حوالي ثلثي المشاريع. كان الاختلاف الرئيسي بين المنطقتين في الاعتبار هو النسبة المئوية للشركات التي أجرت مقابلات مع العملاء: 73٪ في أمريكا الشمالية و 58٪ في أوروبا ، دون اختلافات كبيرة في السلوك بين الشركات الصغيرة والكبيرة الحجم.

فيما يتعلق بمستوى المصطلحات في وثائق المتطلبات ، يجد المرء أن 79٪ من هذه الأخيرة مصاغة بلغة طبيعية (الشكل 5). بالنسبة لتحليل المراسلات ، تم دمج الطريقتين الأخيرتين (اللغة المنظمة والرسمية).

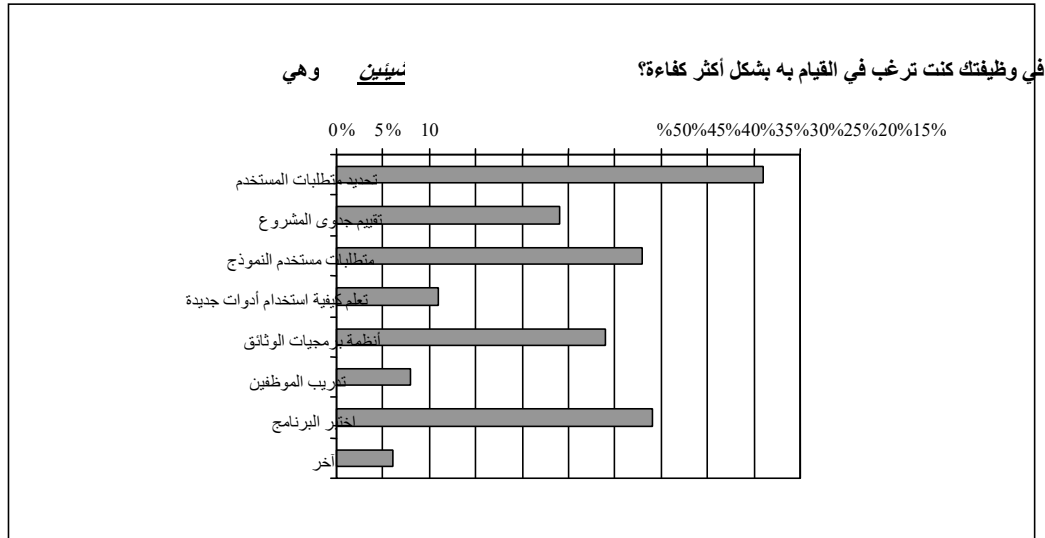


الشكل 5 - مستوى المصطلحات في وثائق المتطلبات

لم يظهر تحليل الترابط بين استخدام اللغة الطبيعية والعوامل الأخرى التي تم فحصها أي ارتباط كبير بنوع الشركة ، ولا باعتماد منهجية.

جانب آخر مهم يتعلق بكل من الطلب المحتمل على أداة CASE القائمة على البرمجة اللغوية العصبية بشكل خاص وتطوير البرمجيات بشكل عام ، هو معرفة المجال المطلوبة لفهم المشكلة بشكل كاف بحيث يمكن تحديد متطلبات المستخدم. في الواقع ، في ظل وجود مستويات عالية من المعرفة المتخصصة ، يجب تكييف الأداة مع احتياجات كل عميل إذا كانت ستعمل بكفاءة في إعدادات الشركات المختلفة. على النقيض من ذلك ، يسمح المستوى المنخفض للغاية بتطوير أداة قياسية واحدة قادرة على العمل في مجالات التطبيق المختلفة. وفي هذا الصدد، تبين أن المستجيبين يحتاجون إلى مستوى متوسط (54٪) إلى مستوى مرتفع (34٪) من المعرفة بالمجال. واتضح أيضا أنه كلما ارتفع مستوى المعرفة بالمجال المطلوب لتطوير البرمجيات، زاد استخدام المنهجيات (9٪ للمستويات المنخفضة، و 53٪ للمستويات المتوسطة، و 38٪ للمستويات العالية) وأدوات تحليل المتطلبات وتصميم المستوى الأعلى (2٪ و 56٪ و 42٪ على التوالي).

[3] فيما يتعلق بكفاءة عمليات الإنتاج ، عند الانتهاء من دراسة السوق ، كان من المهم تحديد أنشطة البرمجيات التي ينظر إليها على أنها حاسمة ، وكذلك وزنها بالنسبة للمتطلبات (السؤال 16).

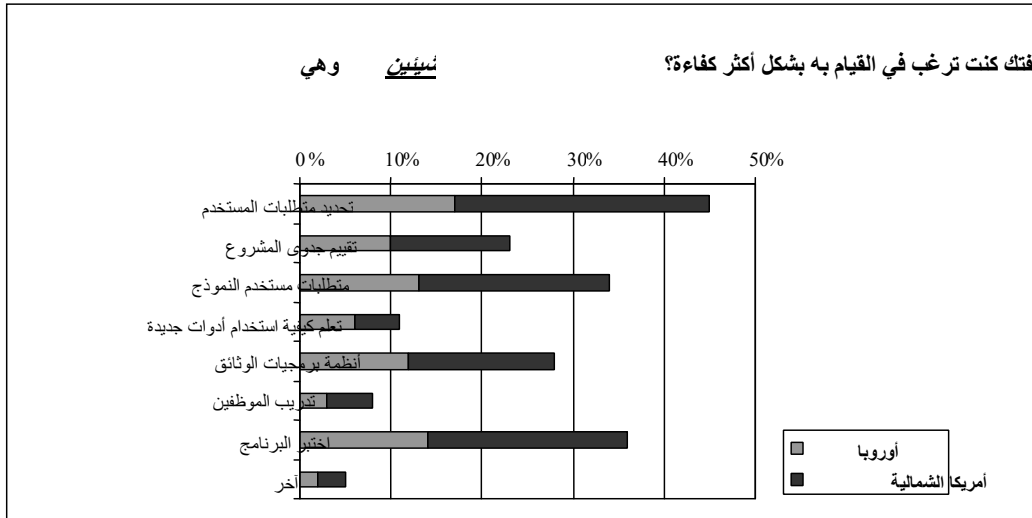


الشكل 6 - الأنشطة التي ينظر إليها على أنها حاسمة في تطوير البرمجيات

عند تفسير الإجابات على هذا السؤال ، تجدر الإشارة إلى أنه تم طلب اختيارين ، وبالتالي الحصول على نتائج أعلى من 100 بالمائة. ويبين الشكل 6 أن نسبة عالية من المجيبين ذكرت "تحديد متطلبات المستخدم" و "متطلبات المستخدم النموذجية" كأولويات. على عكس حالة "تحديد متطلبات المستخدم" - التي كانت مستقلة إلى حد كبير عن اللغة المستخدمة لنمذجة المتطلبات (46% للغة الطبيعية الشائعة ، و 37% للغة الطبيعية المنظمة و 50% للغة الرسمية) و "اختبار البرنامج" (35% ، 32% ، 38% على التوالي) - بالنسبة لـ "متطلبات المستخدم النموذجية" ، كانت النسب المئوية 38% للغة الطبيعية المشتركة و 13% للغة الرسمية ، وفقا للتوقعات. نتيجة أخرى جديرة بالملاحظة هي أن الاختبار كان ينظر إليه على أنه حاسم من قبل نسب أعلى (تتراوح من 19% إلى 46%) من المجيبين الذين لم يستخدموا أي أدوات على الإطلاق. يتم عرض نمط مماثل من خلال مستوى المعرفة بالمجال اللازم ، حيث كان ينظر إلى الاختبار على أنه أكثر أهمية من جميع الأنشطة الأخرى في مستويات المعرفة المنخفضة (63% ، مقارنة بـ 32% و 30% لمستويات المعرفة المتوسطة إلى العالية). ومما يثير الاهتمام أيضا حقيقة أن "تعلم استخدام أداة جديدة" تم اختياره من قبل نسبة أعلى من المجيبين الذين أعلنوا أنهم لم يستخدموا أداة لتحليل المتطلبات مقارنة بأولئك الذين قالوا بدلا من ذلك إنهم استخدموا أداة من هذا النوع.<sup>68</sup>

تتطلب أهمية هذا السؤال مقارنة النتائج لأوروبا وأمريكا الشمالية (انظر الشكل 7). كما تم إجراء تحليل المراسلات - الوارد في الجزء الثاني من هذا القسم - مع مراعاة مركزية هذا السؤال فيما يتعلق بأهداف أبحاث السوق ، حيث تصبح الأنشطة التي تعتبر الأكثر أهمية حاسمة عند تحديد الملامح الشخصية.

<sup>68</sup> تجدر الإشارة إلى أن حوالي ثلث الملاحظات النهائية تتعلق بدور وأهمية المتطلبات. مع الأخذ في الاعتبار الأهداف المختلفة للاستطلاعات الموضحة في [30،31] ، يمكننا مقارنة هذه النتائج مع تلك التي تم الحصول عليها لسؤال من هناك ein على الأهمية النسبية المتصورة لمشاكل البرمجيات في أوروبا (معظم مشاكل البرمجيات هي في مجال مواصفات المتطلبات وإدارة متطلبات العملاء ؛ بعد التوثيق والاختبار) وعلى النطاق المتصور لـ GE نموذج عملية NERIC (تحديد متطلبات النظام ، 78%).  
<sup>69</sup> وفي هذا الصدد، نقبس ملاحظة وردت في أحد الاستبيانات: "أكره أن أكون ساخرا ، لكن لا تكاد توجد أي أدوات جديرة بالاهتمام. النفقات العامة في تعلم استخدامها كبيرة جدا بالنسبة للباييووما يليها".



الشكل 7 - الأنشطة التي ينظر إليها على أنها حاسمة في تطوير البرمجيات (أوروبا مقابل أمريكا الشمالية)

ورداً على سؤال "ما هو الشيء الأكثر فائدة لتحسين الكفاءة اليومية العامة؟" اختارت الأغلبية (64٪) خيار "الأتمة"، في حين تم اختيار "الاستعانة بمصادر خارجية" بنسبة 7٪ و "التفويض الداخلي" بنسبة 29٪. وخلافاً للتوقعات، لم تظهر فروق معينة بين الردود على هذا السؤال فيما يتعلق بحجم الشركة، حيث كان الاختلاف الوحيد المهم يتعلق بالشركات التي يتراوح عدد موظفيها بين 6 و 20 موظفاً، حيث كانت النسبة المئوية لاختيار "التفويض الداخلي" ضعف النسبة المئوية لمجموعات الشركات الأخرى، وهو اختلاف قد يكون راجعاً إلى أوجه القصور التنظيمية. ومن المثير للاهتمام أن النسبة المئوية للمستجيبين الذين استخدموا منهجية أو أداة تحليل المتطلبات واعتقدوا أنه من الأقل أهمية زيادة مستوى التفويض الداخلي كانت أعلى من متوسط العينة بأكملها. وبدلاً من ذلك، لم تكن هناك اختلافات فيما يتعلق بالوثائق المتاحة لتحليل الاحتياجات.

يظهر التحليل المشترك للسؤالين حول كفاءة عمليات إنتاج البرمجيات أن نسبة أكبر من المستجيبين الذين يعتقدون أنه من المهم زيادة مستوى الأتمة قد اختاروا سابقاً "تعلم استخدام أداة جديدة" و "متطلبات المستخدم النموذجية" (الجدول 4).

الجدول 4 - كفاءة عمليات تطوير البرمجيات

ما هو الشيء الأكثر فائدة لتحسين الكفاءة اليومية العامة؟			ما هما الأمران في وظيفتك اللذان ترغب في القيام بهما بشكل أكثر كفاءة؟
التفويض الداخلي	الاستعانة بمصادر خارجية	اتمته	
22%	9%	69%	تحديد متطلبات المستخدم
44%	12%	44%	تقييم جدوى المشروع
21%	4%	75%	متطلبات مستخدمي النموذج
14%	0%	86%	تعلم كيفية استخدام أداة جديدة
24%	5%	71%	أنظمة برمجيات الوثائق

تدريب الموظفين	18%	0%	82%
اختبر البرنامج	67%	4%	29%
آخر	43%	14%	43%

بالنسبة للسؤال الأخير ، فيما يتعلق بمتوسط التأخير في تسليم البرنامج ، تم تحقيق أفضل أداء من قبل الشركات التي لديها 6-20 موظفا (29٪ منهم تم تسليمهم بأقل من أسبوع واحد من التأخير و 59٪ بأقل من شهر واحد) وأولئك الذين باعوا مباشرة إلى المستهلك النهائي (ربما لأسباب تعاقدية). وإن لم يكن ذلك إلى حد ذي دلالة إحصائية ، فإن الشركات التي تستخدم لغة رسمية قدمت بأقل تأخير ، على الرغم من عدم وجود اختلافات جوهرية فيما يتعلق بالتأخير لأكثر من شهر واحد (26٪ للغة الطبيعية الشائعة ، و 33٪ للغة الطبيعية المنظمة ، و 25٪ للغة الرسمية). يتطلب التفسير العادل لهذه النتائج أن يتذكر المرء أن الإجابات لا تأخذ في الاعتبار طول المشاريع. ومع ذلك ، بافتراض أن متوسط التأخير أقل من أسبوع واحد يتوافق مع الشركات التي تقدم البرنامج في المتوسط خلال الوقت المحدد ، تم الإبلاغ عن نتائج مماثلة في [32] ، حيث ذكر أكثر من 80٪ من المستجيبين أن مشاريعهم كانت متأخرة في بعض الأحيان أو عادة.

بالنظر إلى الغرض من هذه الدراسة ، ولا سيما مسألة ما إذا كان هناك سوق لأداة CASE القائمة على البرمجة اللغوية العصبية لتحليل المتطلبات ، تؤكد النتائج المقدمة حتى الآن تصور تحليل المتطلبات على أنه أمر حاسم لتطوير الأنظمة ، والاستخدام الواسع النطاق للنهج الموجه للكائنات و UML ، والدور الهام للغة الطبيعية. تحديداً:

- تتبنى أكثر من 80٪ من الشركات منهجية لتطوير برامجها ، ويتبنى ما يقرب من 68٪ منها طريقة موجهة للكائنات (UML أو إحدى الطرق المدمجة في UML).
- غالبية الوثائق المتاحة لتحليل المتطلبات هي باللغة الطبيعية ويتم توفيرها إما من قبل العميل أو الحصول عليها عن طريق المقابلات.
- معرفة المجال المطلوبة متوسطة إلى عالية.
- يتم استخدام الأدوات التي تدعم تحليل المتطلبات وتصميم المستوى الأعلى في أقل من ثلث الحالات.
- ومع ذلك ، ينظر إلى متطلبات تحديد ونمذجة على أنها لا تقل أهمية عن اختبار البرنامج.
- يشير حوالي 64٪ من المستجيبين إلى مستوى أعلى من الأتمتة باعتباره الوسيلة الأكثر فائدة لتحسين الكفاءة اليومية.

تعمل كل هذه العناصر معا لتأكيد وجود طلب محتمل على

أداة CASE تعتمد على البرمجة اللغوية العصبية. لتبرير هذا الادعاء ، أجرينا دراسة تحليل المراسلات (CA). وهذا يعني استخدام تقنية إحصائية مناسبة لدراسة العلاقات بين الطرائق مع متغيرين أو أكثر يمكن تمييزهما ، وعادة ما تكون نوعية. يتم وصف الخطوات الرئيسية لتحليل المراسلات بإيجاز على النحو التالي:

- (5) تحديد سحابة من النقاط (صفوف وأعمدة جدول الطوارئ) في فضاء متجه متعدد الأبعاد ؛
- (6) اختر الهيكل المتري في هذه المساحة ؛
- (7) إنتاج ملاءمة السحابة في 1) إلى مساحة فرعية متغيرة منخفضة الأبعاد يتم عرض النقاط عليها (ملفات تعريف الصفوف والأعمدة) ؛

(8) إعطاء تفسير لمجموعات النقاط المقابلة لتوقعات الصفوف والأعمدة في جدول الطوارئ الأصلي ؛ تحليل مساهماتها المطلقة كأدلة لتفسير الأبعاد الأساسية ومساهماتها النسبية (ما يسمى بالارتباطات التربيعية) للإشارة إلى مدى جودة وصف النقاط على طول البعد المدروس.

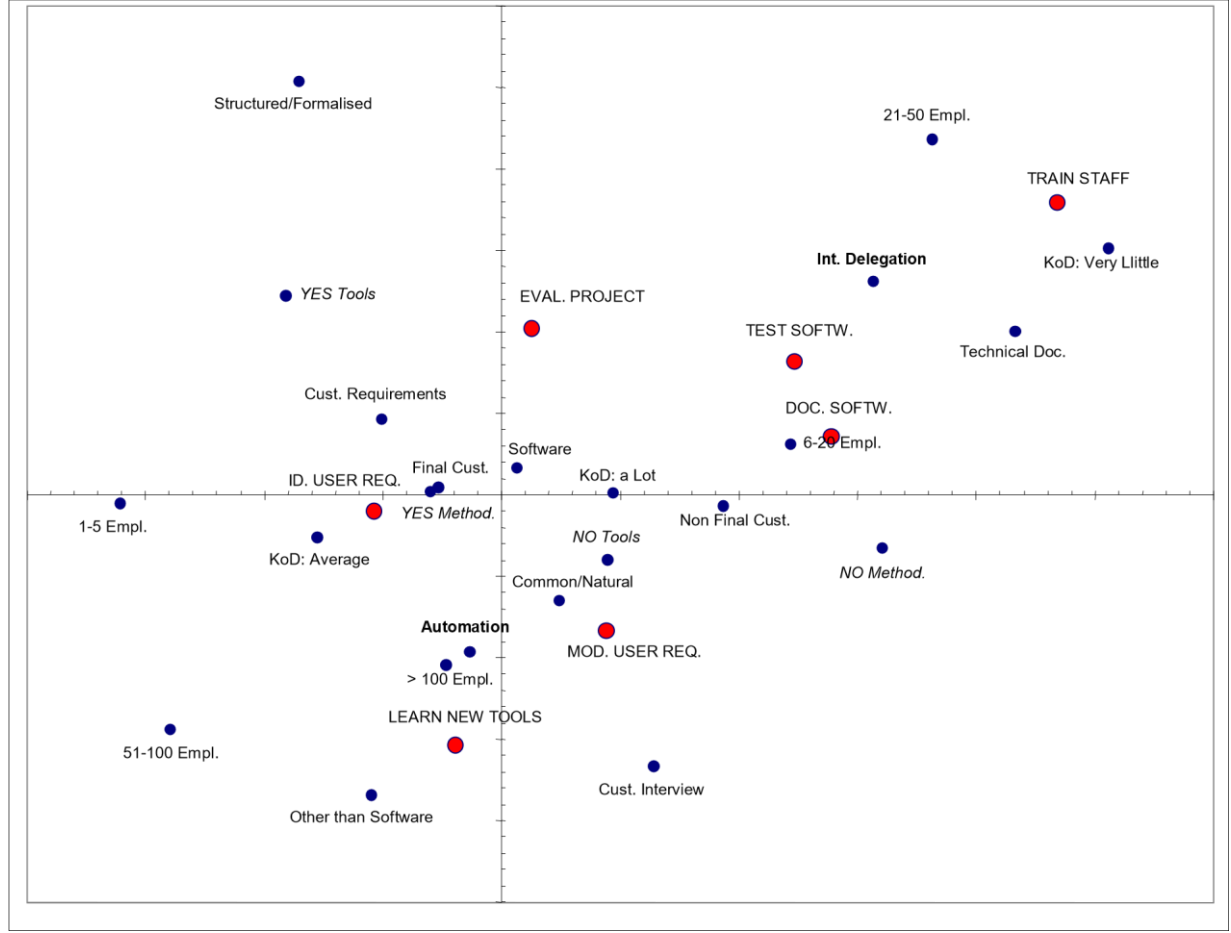
تشبه هندسة CA إلى حد كبير الوصف الهندسي لكارل بيرسون [33] لتحليل المكونات الرئيسية. يتم تعريف قرب النقاط من خط أو مستوى أو بشكل عام إلى مساحة فرعية منخفضة الأبعاد ، على أنها مجموع المسافات التربيعية من النقاط إلى الفضاء الفرعي. بشكل عام ، من المهم تجنب المقارنة المباشرة للمسافات بين إسقاطات ملفات تعريف الصفوف والأعمدة لأنها تنتمي إلى مساحات فرعية مختلفة منخفضة الأبعاد وقد يؤدي التفسير الخام لمسافاتها إلى استنتاجات مضللة.

هنا نظرنا في المرجع المصدق الذي يتضمن أحد بنود الاستبيان (ما يجب القيام به بشكل أكثر كفاءة) كمتغير تابع وبعض المتغيرات الأخرى التي تم جمعها (عدد الموظفين ، الأعمال الأساسية ، نوع البرامج المنتجة ، استخدام أي منهجية ، بدء التوثيق ، مستوى المصطلحات ، استخدام أي أداة ، معرفة المجال ، شيء لتحسين الكفاءة اليومية ، متوسط التأخير في تسليم البرمجيات) كمتغيرات مستقلة من أجل التحقق مما إذا كانت الإجابة على هذا البند تتأثر بطرائق المتغيرات الأخرى ومدى تأثيرها وتحديد بعض التجميعات ذات الصلة من الطرائق التي يمكن أن تكشف عن طلب السوق المحتمل على أداة CASE القائمة على البرمجة اللغوية العصبية.

نقدم هنا نتيجة تطبيق CA بناء على الإجابات على السؤال المتعلق بالأنشطة التي تعتبر الأكثر أهمية (انظر الشكل 8).<sup>70</sup>

يمكن الوصول إلى تفسير أولي للرسم البياني من خلال النظر إلى المحاور. على وجه التحديد ، يمكن للمرء أن يفسر المحور الرأسي من الناحية التنظيمية ، على افتراض أن طلب المزيد من الأتمتة بدلا من التفويض الداخلي يرجع إلى هيكل تنظيمي قوي إلى حد ما. وفي الوقت نفسه ، يتوافق المحور الأفقي مع هندسة أو نهج غير رسمي لتطوير البرمجيات اعتمادا على استخدام أو عدم استخدام المنهجيات والأدوات لدعم التحليل والتصميم.

<sup>70</sup> جدول الطوارئ متاح في <http://on-line.cs.unitn.it>.



شكل 8 - ناتج تحليل المراسلات<sup>71</sup>

وفقا لهذا التفسير للرسم البياني ، هناك نوعان من منافذ السوق المحتملة.

يتوافق مكان السوق الأول مع الشركات التي تتبنى منهجيات وأدوات لدعم تحليل المتطلبات والتصميم عالي المستوى. يمكننا أن نفترض بأمان أنهم يستخدمون عملية تطوير برمجيات "صناعية" بدلا من "حرفية". بالنسبة لهذا النوع من الشركات ، يعتبر تقييم المشروع نشاطا مهما ، إلى جانب تحديد المتطلبات. هذان النشاطان ، من بين الأنشطة الممكنة المدرجة في الاستبيان ، هما الأكثر تعددا للتخصصات وفي نفس الوقت الأكثر صعوبة في التنظيم. على وجه الخصوص ، لأغراض دراستنا ، يمكن دعم تحديد المتطلبات بكفاءة من خلال أدوات قادرة على تحليل المستندات بلغة طبيعية. علاوة على ذلك ، بالنسبة لهذا النوع من الشركات ، يجب أن تكون الأداة متخصصة للحصول على مستوى مناسب من المعرفة بالمجال لمجال معين من تطوير البرمجيات. يوفر العميل مستندات المتطلبات ويتم تسليم البرنامج المنتج بدوره إلى العميل. بالنسبة للنهج الموجه نحو العملاء ، فإن هذا يعني وجود إمكانية محدودة فقط لمطالبة العميل بكتابة المستندات في شكل مقيد من اللغة الطبيعية ؛ ومع ذلك ، تتلقى هذه الشركات أحيانا المستندات في شكل منظم إلى حد ما (رسمي). في هذه الحالات ، من الممكن تصور استخدام تقنيات لغوية أقل تطورا لتحليل وثائق المتطلبات من أجل إنتاج نماذج مفاهيمية باستخدام النهج الموجه للكائنات.

<sup>71</sup> ولم يتم تمثيل نقطتين ("أخرى" في السؤال 16 فيما يتعلق بالأنشطة الحرجة، و "الاستعانة بمصادر خارجية" بالنسبة للسؤال 17) بسبب بعدهما الكبير عن مركز (منخفض التردد) ، مما يجعل الرسم البياني أكثر قابلية للفهم.



يشمل مجال السوق الثاني الشركات المتوسطة أو الكبيرة الحجم التي لا تستخدم منهجيات ولا أدوات لدعم تحليل المتطلبات والتصميم عالي المستوى. ومع ذلك ، فإنهم يرون أن نمذجة المتطلبات أمر بالغ الأهمية ، إلى جانب أنشطة أخرى مثل توثيق البرامج واختبارها ، والتي يتم دعمها بالفعل بطرق مختلفة بواسطة أدوات CASE الحالية. يمكن للمرء أن يستنتج بشكل معقول أن هذه المجموعة الثانية من الشركات تشكل أيضا مكانا مناسباً في السوق لأداة CASE التي تتيحها الأدوات اللغوية. في الواقع ، يمكن لـ CASE من هذا النوع دمج وظائف CASE التقليدية ، مما يفضل اعتماد نهج هندسي في تطوير البرمجيات. نشاط آخر يعتبر حاسماً هو تعلم أدوات جديدة ، وهي عقبة يمكن التغلب عليها من خلال اعتماد CASE الذي يستخدم اللغة الطبيعية على نطاق واسع. إن الإشارة إلى نمذجة المتطلبات بدلاً من تحديدها تسلط الضوء على حقيقة أن مشكلة على مستوى مواصفات المتطلبات يمكن أن تخفي مشاكل أعمق تتعلق باستنباط المتطلبات (يمكن دعمها بواسطة أنظمة التعرف على الكلام وجميع الوظائف المتوخاة في النقطة / من القسم 2). وهذا ما تؤكده إلى حد ما حقيقة أن تحديد المتطلبات بدلاً من نمذجتها يعتبر أمراً بالغ الأهمية من قبل الشركات التي تعتمد نهجاً أكثر تنظيماً لتطوير البرمجيات.

أحد الجوانب المهمة لهذا البحث هو التطبيق الأوسع للنتائج. كما لوحظ ، فإن هذا البحث وصفي ، يعتمد على عدد كبير من الاستبيانات (من بين أعلى الاستبيانات التي رأيناها في دراستنا) ، ولكنه لا يمثل السكان بشكل كامل. والحقيقة هي أنه بالنسبة لصناعة البرمجيات ، لا توجد ببساطة معلومات كافية عن السكان المرجعيين للسماح بتمديد النتائج بشكل هادف وصحيح إحصائياً.<sup>72</sup>

بعد قلبي هذا ، نؤكد أنه من المفيد إجراء مقارنة مع البيانات المتاحة في الأدبيات. ويخلص الجدول التالي أهمها (الجدول 5). تجدر الإشارة إلى ندرة البيانات الموجودة. على الرغم من أن الدراسات الاستقصائية التي تشير إليها هذه النتائج مختلفة تماماً ، إلا أن أوجه التشابه بينها تبرز.<sup>74,73</sup>

الجدول 5 - مقارنة مع النتائج المتعلقة بالدراسات الاستقصائية الأخرى وسوق CASE

حالة قائمة على البرمجة اللغوية العصبية أداة السوق عبر الإنترنت الأبحاث - 1999 (142 مؤسسات)	حالة الممارسة مسح حول RE - 1999 <sup>41</sup> (12 مؤسسات)	تطوير المشاريع - حالة الممارسة - 1997 <sup>42</sup> (78 مؤسسات)	الحصة السوقية OO CASE Tools - 1998 <sup>43</sup>
بيع للمستخدم النهائي	84%	83%	-
SW كعمل أساسي	82%	66%	-
استخدم طرق OO	68%	50%	39% (استخدم اتجاه O)

<sup>72</sup> الاستثناءات البارزة هي الدراسات الاستقصائية التي أجراها المعهد الأوروبي للبرمجيات: <http://www.esi.es>.

<sup>73</sup> تم إجراء هذه المسوحات بأهداف مختلفة وباستخدام طرق مختلفة و sa ممبلس. استخدم المسح الموصوف في [25] 78 استبياناً تم تجميعها بشكل أساسي من قبل مديري أو مديري تطوير نظم المعلومات في الشركات العاملة خارج مجال البرمجيات ، بينما يبلغ الاستبيان الفنلندي عن نتائج تتعلق بـ 12 شركة فنلندية، 8 منها عملت حصرياً في مجال البرمجيات.

<sup>74</sup> انظر [23].<sup>42</sup>

انظر [25]. لاحظ أنه عندما تم إجراء هذا الاستطلاع ، تم اعتماد UML للتو كميّار من قبل OMG.

	استخدم منهجية دورة الحياة الرسمية		%77 %79 %30 ( & أدوات تصميم المستوى الأعلى) %52 %46	استخدام UML متطلبات اللغة الطبيعية استخدم أدوات RA استخدم الوردة العقلانية تحديد متطلبات المستخدم
< 48%	-	-		
	-	100%		
	29% (استخدم أداة CASE الأمامية)	0%		
33%	-	-		
	-	66%		

يمكننا أيضا أن نستشهد هنا ببعض البيانات الموجودة في [34] ، والتي تحتوي على مؤشرات مفصلة للنسبة المئوية للصفحات باللغة الطبيعية أو أشكال مماثلة - نص يحتوي على كلمات رئيسية وتعداد هرمي وجدول - لثلاثة مشاريع ، لها قيم تتراوح من 82٪ إلى 99٪ (73٪ و 43.9٪ و 34.4٪ على التوالي ، فقط لنص اللغة الطبيعية).

ومن الجوانب الأخرى التي تمكن من إجراء تقييم إيجابي لنتائج الدراسة الاستقصائية انخفاض النسبة المئوية لعدم الردود (1.65 في المائة) وحقيقة أنه في حالة الردود التي تم اختيار خيار "أخرى" لها، تم تحديد مواصفات في 91٪ من الحالات.

## 5. الاستنتاجات

نظرا لأن الهدف الرئيسي لهذا المشروع البحثي كان تقييم ما إذا كان هناك سوق لأدوات CASE التي تدعم البرمجة اللغوية العصبية ، فإن أهم نتيجة هي أن غالبية المستندات المتاحة لتحليل المتطلبات يتم توفيرها من قبل العميل ومصاغة بلغة طبيعية "حقيقية" ، مما يؤدي إلى استنتاج مفاده أن استخدام التقنيات والأدوات اللغوية قد يؤدي دورا حاسما في توفير الدعم لتحليل المتطلبات.

نظرا لأن النهج الهندسي يقترح استخدام أدوات لغوية مناسبة للغة المستخدمة في الوصف السردى لمتطلبات المستخدم ، نجد أنه في معظم الحالات يكون من الضروري استخدام أنظمة البرمجة اللغوية العصبية القادرة على تحليل المستندات بلغة طبيعية كاملة. إذا تم التحكم في اللغة المستخدمة في المستندات (إعطاء مجموعة فرعية من اللغة الطبيعية) ، فمن الممكن استخدام أدوات لغوية أبسط وبالتالي أقل تكلفة ، والتي تكون متاحة بالفعل في بعض الحالات. يمكن أيضا استخدام أدوات من هذا النوع لتحليل المستندات بلغة طبيعية كاملة ، حتى لو كانت هناك حاجة في هذه الحالة إلى مزيد من التشاور مع المحللين لتقليل تعقيد اللغة المستخدمة في مستندات الإدخال أو للتدخل تلقائيا في النماذج المنتجة كمخرجات. علاوة على ذلك ، هناك حاجة في كثير من الحالات ، إلى جانب التمثيل الكافي للمعرفة المشتركة / المشتركة ، إلى المعرفة المتخصصة بالمجال. مرة أخرى ، تتطلب إدارة معرفة الخبراء استثمارات أكبر لتكييف الأداة مع احتياجات الشركة.

أما بالنسبة للطلب المحتمل على أدوات CASE القائمة على البرمجة اللغوية العصبية ، فقد تم تحديد ملفين شخصيين للشركة ، يقابلان مكانتين متميزتين في السوق. الأول يتألف من الشركات التي لديها نهج هندسي لتطوير البرمجيات والتي أشارت - من بين النشاطين المرتبطين بتحليل المتطلبات - إلى تحديد المتطلبات باعتبارها الأكثر أهمية. في هذه الحالة ، يمكن تكوين الأداة كوحدة نمطية للتكامل مع أداة CASE

المستخدمة بالفعل من قبل الشركة ، وستوفر الدعم للمراحل التي تكون فيها الأدوات الحالية غير كافية. في مكانة السوق الثانية ، تستخدم تقنيات اللغة الطبيعية لتسهيل اعتماد أداة CASE وبشكل أعم "أفضل الممارسات" لتطوير البرمجيات ، بالنظر إلى أنه إلى جانب نمذجة المتطلبات ، أشارت هذه الشركات أيضا إلى الأنشطة الحاسمة التي تم فيها تطوير مساهمة هندسة البرمجيات بشكل جيد (الاختبار أو توثيق البرامج ، على سبيل المثال).

يمكننا أيضا تقديم بعض الملاحظات الأولية هنا فيما يتعلق بالميزات المتوقعة من أداة تعتمد على البرمجة اللغوية العصبية ، انطلاقا من المقابلات مع محلي / مهندسي الأنظمة ومديري المشاريع في كل من الشركات الصغيرة والمتوسطة الحجم. على وجه التحديد ، يؤكدون الافتراضات التي تم إجراؤها فيما يتعلق بالطلب المحتمل والاهتمام بالميزات التالية:

- إمكانية تسريع إنتاج نماذج التحليل وإنشاء نماذج بسرعة لاستخدامها في التفاعلات مع المستخدمين وفي مجموعات المشاريع. وحقيقة أن نماذج الطبقات، على سبيل المثال، قد تحتوي على فئات زائفة أو أن بعض الفئات قد تكون مفقودة اعتبرت أقل أهمية إذا أنتجت النماذج تلقائيا.

(43) بيانات مؤسسة البيانات الدولية.

- اعتبرت الأداة مفيدة أيضا لتدريب المحللين، مع عرض النصوص والنماذج المقابلة، سواء بالنسبة للمحللين المبتدئين أو لإعادة تدريب أولئك الذين ليسوا على دراية بالنهج الموجه نحو الكائنات (يبدو أن المشكلة الأخيرة أكثر أهمية بالنسبة للشركات الصغيرة الحجم).

- إمكانية دمج الأداة مع أدوات CASE لرسم المخططات باستخدام العناصر التي حددتها الخوارزمية واستخدام أدوات لإدارة الوثائق.

أخيرا ، بالنسبة لبعض الأسئلة في الاستطلاع (على سبيل المثال ، استخدام المنهجيات ونماذج E-R ، واستخدام أدوات الدعم في المراحل الأولية من التطوير) ، فإن المساهمات التي تقدمها هذه الورقة في هذا المجال تتجاوز حدود أبحاث السوق كما هو موضح في العنوان. وأكد بعض التوقعات (انتشار النهج الموجه للكائنات) ، والتي يمكن أن تبدو واضحة على السطح ، لكنها لم تكن مدعومة بشكل كاف بالبيانات الثابتة. كما أكدت وجود إمكانيات كبيرة لاعتماد أدوات وأساليب هندسة البرمجيات [35].

## مراجع

- [4] Franch M ، Mich L ، Osti L. البحث عبر الإنترنت كأداة قرار للتسويق و استراتيجيات الإدارة. في Proc. تكنولوجيا المعلومات لإدارة الأعمال - 16th ITBM2000 ، IFIP WCC ، بكين ، الصين ، 25-21 أغسطس 2000 ، Gan R (ed) ، بكين ، 2000 ، ص 743-737.
- [5] D'Elia M. أبحاث السوق عبر الإنترنت: تطبيق على مجال البرمجيات. أطروحة للحصول على درجة علمية، جامعة تريننتو (باللغة الإيطالية)، 2000.
- [6] Loucopoulos P، Karakostas V. هندسة متطلبات النظام. ماكجرو هيل 1995 [4] تشيوتشيتي ن ، ميتش ل. سوق أدوات CASE الموجهة للكائنات. تقرير تقني ، قسم علوم الكمبيوتر والإدارة ، جامعة تريننتو (باللغة الإيطالية) ، 31 ، 2000.
- [5] Burg J.F.M. أداة لغوية في هندسة المتطلبات ، IOS ، أمستردام ، 1997. [6] ريان ك. دور اللغة الطبيعية في هندسة المتطلبات. IEEE 1992 ؛ 240242.
- [36] تشن PP-S. بنية الجملة الإنجليزية ومخططات علاقات الكيان. علوم المعلومات، 1983؛ 29: 149-127.

- [37] أمبريولا الخامس، جيرفاسي ضد. بيئة للبناء التعاوني لقواعد متطلبات اللغة الطبيعية. في Proc.8<sup>th</sup> ICRE ، مطبعة جمعية الكمبيوتر 1999 ، IEEE ، ص 124-130.
- [38] جورستون، مورينو إيه إم، لوبيز م. كيفية استخدام الأدوات اللغوية لتحليل IEEE SW OO. ، مايو / يونيو 2000 ، 80-89.
- [39] فوكس إن إي ، شويتر آر أتويمبتو الإنجليزية المسيطر عليها. في: CLAW'96 ، ورشة عمل دولية 1st حول تطبيقات اللغة الخاضعة للرقابة ، جامعة كاثوليكي ، لوفين - بلجيكا 1996.
- [40] Delisle S، Barker K، Biskri I. التحليل الموجه للكائنات: الحصول على المساعدة من الأدوات اللغوية الحاسوبية القوية. في G Friedl، HC Mayr (محرران) تطبيق اللغة الطبيعية على نظم المعلومات، OCG، الصفحات 167-172، 1999.
- [41] ميتش ل ، جاريليانو ر. مقاييس الغموض في هندسة المتطلبات. في بروك. ICS2000 الدورة السادسة عشرة لمجلس الكنائس الدولي، بيجين، الصين، 21-25 آب/أغسطس 2000، الصفحات 39-48.
- [42] ديفيس إي إم. الانسجام في التجمعات. برنامج IEEE ، مارس / أبريل 1998 ، ص 6-8.
- [43] Nitto E Di ، Fuggetta A. التغيير مقابل التوحيد: تحد لمنظمات تطوير SW. ريفيستا دي المعلوماتية، 25 ، AICA 1995، (4): 267-279.
- [44] Mylopoulos J. نمذجة المعلومات في زمن الثورة ، نظم المعلومات ، مايو 1998 ؛ 23 (3-4): 127-156.
- [45] Rugg G ، Hooper S. معرفة ما لا يمكن معرفته: أسباب وطبيعة المتطلبات المتغيرة. في J Eder، N Maiden، M Missikoff (eds) Proc. 1st Int البندقية ، 25-27 سبتمبر 1999 ، ص 183-192.
- [46] مؤتمر فهم رسالة إجراءات MUC-3 ، MUC-4 ، MUC-5 ، MUC-6 ، مورغان كوفمان 1991 ، 1992 ، 1993 ، 1995 ، 1998 ، [http://www.itl.nist.gov/iaui/894.02/related\\_projects/muc/index.html](http://www.itl.nist.gov/iaui/894.02/related_projects/muc/index.html)
- [47] فابريني إف ، فوساني إم ، جيرفاسي الخامس ، جنيسي إس ، روجيري س. تحقيق الجودة في الطبيعة متطلبات اللغة. في: Proc Int SW Quality Week ، S.Francisco CA ، مايو 1998
- [48] لايتبرغ أو ، أتكينسون سي ، شليش إم ، إل إمام ك. مقارنة تجريبية لتقنيات القراءة للكشف عن العيوب في وثائق تصميم ل UML. من S &SW ، شمال هولندا - السفير 2000 ؛ 53: 183-204.
- [49] Canzano G. معالجة اللغة الطبيعية في أبحاث السوق: التحليل التلقائي للردود على الأسئلة المفتوحة. أطروحة جامعية، جامعة ترينتو (باللغة الإيطالية)، 1999.
- [50] Mich L. NL-OOPS: من اللغة الطبيعية إلى متطلبات OO باستخدام نظام معالجة اللغة الطبيعية LOLITA. في: J من هندسة اللغة الطبيعية ، مطبعة جامعة كامبريدج 1996 ؛ 2 (2): 161-187.
- [51] ميتش ل ، جاريليانو ر. مشروع NL-OOPS: نمذجة OO باستخدام NLPS LOLITA. في Proc. 4<sup>th</sup> Int. Conf. NLDB'99 ، كلاغنفورت ، 17-19 يونيو 1999 ، Friedl G ، Mayr HC (eds) ، تطبيق اللغة الطبيعية على نظم المعلومات ، فيينا 1999 ، 215-218.
- [52] نيكولا يو، ساجانيمي ج، كيلفيانين ه. مسح حالة الممارسة حول هندسة المتطلبات في الشركات الصغيرة والمتوسطة الحجم. تقرير البحث 1. جامعة لايبزانتا للتكنولوجيا ، 2000.
- [53] Zvegintzov N. الأسئلة التي يتم التسول فيها بشكل متكرر وكيفية الإجابة عليها. IEEE SW مارس / أبريل 1998 ؛ 93-96.
- [54] جلاس آر ، هوارد أ. تطوير البرمجيات حالة من الممارسة. إدارة تطوير النظام ، يونيو 1998 ، 7-8.

- [55] كوملي ب. استخدام الإنترنت كطريقة لجمع البيانات. SGA أبحاث السوق ، 1996.
- [56] ويلرايت إس سي ، ماكريدكيس س. طرق التنبؤ. جون وايلي وأولاده ، نيويورك 1985.
- [57] جريناكري جي إم. نظرية وتطبيق تحليل المراسلات. المطبعة الأكاديمية، نيويورك 1984.
- [58] Dutta S، Lee M، Wassenhove L Van. هندسة البرمجيات في أوروبا: دراسة لأفضل الممارسات. IEEE SW مايو - يونيو ، 82-90 ، 1999.
- [59] ESI ، ESPITI - تحليل مسح المستخدم الأوروبي. المعهد الأوروبي للبرمجيات، إسبانيا، تشرين الثاني/نوفمبر 1996.
- [60] ESI ، هندسة النظم في أوروبا. المسح: ملخص النتائج. المعهد الأوروبي للبرمجيات، إسبانيا، آب/أغسطس 1998.
- [61] جينوتشتن إم فان. لماذا تأخر البرنامج؟ دراسة تجريبية لأسباب التأخير في تطوير البرمجيات. IEEE Trans على SWE، يونيو 1991؛ 17 (6): 582-590.
- [62] بيرسون ك. على خطوط وطائرات أقرب ملائمة لأنظمة النقاط في الفضاء. المجلة الفلسفية ، 1901 ، سير. 6 (2): 559-572.
- [63] Melchisedech R. التحقيق في وثائق المتطلبات المكتوبة باللغة الطبيعية ، هندسة المتطلبات. سيرينغر فيرلاغ ، 1998 ، 3: 97-91.
- [64] ESI ، استبيان أفضل ممارسات البرمجيات ، تحليل النتائج. المعهد الأوروبي للبرمجيات، إسبانيا، كانون الأول/ديسمبر 1997.

## الارقام

- شكل 1 - بنية نظام البرمجة اللغوية العصبية للأغراض العامة
- الشكل 2 - عملية توليد النماذج
- الشكل 3 - المحببون حسب منطقة الإقامة الجغرافية
- الشكل 4 - نوع البرنامج
- الشكل 5 - مستوى المصطلحات في وثائق المتطلبات
- شكل 6 - الأنشطة التي تعتبر حاسمة في تطوير البرمجيات
- شكل 7 - الأنشطة التي ينظر إليها على أنها حاسمة في تطوير البرمجيات (أوروبا مقابل الشمال أمريكا)
- شكل 8 - ناتج تحليل المراسلات
- المناضد

- الجدول 1 - حجم الشركة
- الجدول 2 - استخدام أدوات تحليل المتطلبات وتصميم المستوى الأعلى حسب حجم الشركة
- الجدول 3 - مخططات العلاقة بين الكيان والطرق الموجهة للكائنات
- الجدول 4 - كفاءة عمليات تطوير البرمجيات

الجدول 5 - مقارنة مع النتائج المتعلقة بالدراسات الاستقصائية الأخرى وسوق CASE

المواد على الخط (<http://on-line.cs.unitn.it>)

استبيان (نموذج html) قائمة مجموعات

الأخبار التي تم الاتصال بها

رسائل البريد الإلكتروني

تحليل المراسلات

## نموذج للإستبيان

1. كم عدد الموظفين والاستشاريين في شركتك؟

- 5-1
- 20-6
- 50-21
- 100-51
- أكثر من 100

2. ما هو العمل الأساسي لشركتك؟

- البرمجيات
- مواقع الويب
- أخرى: \_\_\_\_

3. ما نوع البرامج التي تطورها شركتك حالياً؟

- برامج الأغراض العامة
- برامج الشبكة
- البرامج الصناعية
- برنامج تطبيق لمنافذ السوق
- أخرى: \_\_\_\_

4. هل تباع شركتك عادة منتجاتها ...

- إلى العميل النهائي
- إلى شركة برمجيات
- إلى متاجر البرمجيات
- أخرى
- ما هو دورك السائد حالياً في الشركة؟
- محلل
- مصمم
- مبرمج
- مهندس نظام / مهندس معماري
- مدير مشروع
- أخرى: \_\_\_\_

