Information Technology for Automatic Identification of Bank Payments for Contracts with Clients

Chief Assist. Prof. Monika Tzaneva, Ph.D.
UNWE, Department "Informatics"
tel.: 0359898766289
e-mail: monika_tzaneva@yahoo.com

Summary: This article presents the development of a complete information technology for automatic identification of bank payments for contracts with clients, whose obligations occur and are paid periodically. Main information problems, which exist during automatic identification of payments, are discussed. A physical data model, which forms the base of this technology, is designed. An algorithm for reliable payment identification based on the analysis of payment description, filled in by clients and the data about contracts, available in core Management Information System is created. A procedure for payment identification, implementing this algorithm and saving the results of its own execution in a form, usable during consequent manual control of these results and during further identification of unidentified payments, is developed. As a conclusion, the achieved results of the implementation of this information technology are presented – significant cutting of the percent of unidentified payments (from 12 to 5 %) has been effected, in conjunction with more than four times reduction of the time, needed for succeeding control and manual identification.

Key words: automatic identification, payments, technology, saving results.

1. Information problems in processing payments under contracts with customers, received in bank transfers

"Information technologies pave the way for business development" [5], this is the shortest definition of the relationship between the information technologies and the business. For this purpose, information technologies must provide the business safe and high quality information services and continuously develop and expand these services to business processes.

Many companies that contract for customer financing in any forms – for example, providing consumer credits in any form (targeted funding, consumer finance, leasing), issuing credit cards or performing various services (public utilities, telephone, Internet, satellite or digital TV)which are paid periodically by bank transfers (incl. electronic), daily handling large volumes of data received payments. In general, the business process processing these payments includes (Figure 1).

1. Daily (or several times throughout the day) servicing banks export files with the encased or transferred amounts in the accounts of the company. These files have different formats, but in all cases the format includes date, amount and terms of payment.

2. Exported files with payments are imported in the database of the company-recipient's business information system.

3. Imported payments should be linked (to be identified) to the contract in which the obligations are paid off. It should be borne in mind that companies of this kind serve between several hundred thousands and several millions customers, as each customer may pay off obligations in many contracts simultaneously.

Obviously, the manual identification of bank payments of this size is unachievable task and requires this procedure to be automated in maximum extent possible. In cases where a client has an individual account, in which the necessary amounts are transferred the identification of payments is a routine procedure. Usually this is the practice in servicing credit cards. In other cases, identification of payments must be made on the basis of a filled in terms of payment and the amount paid.
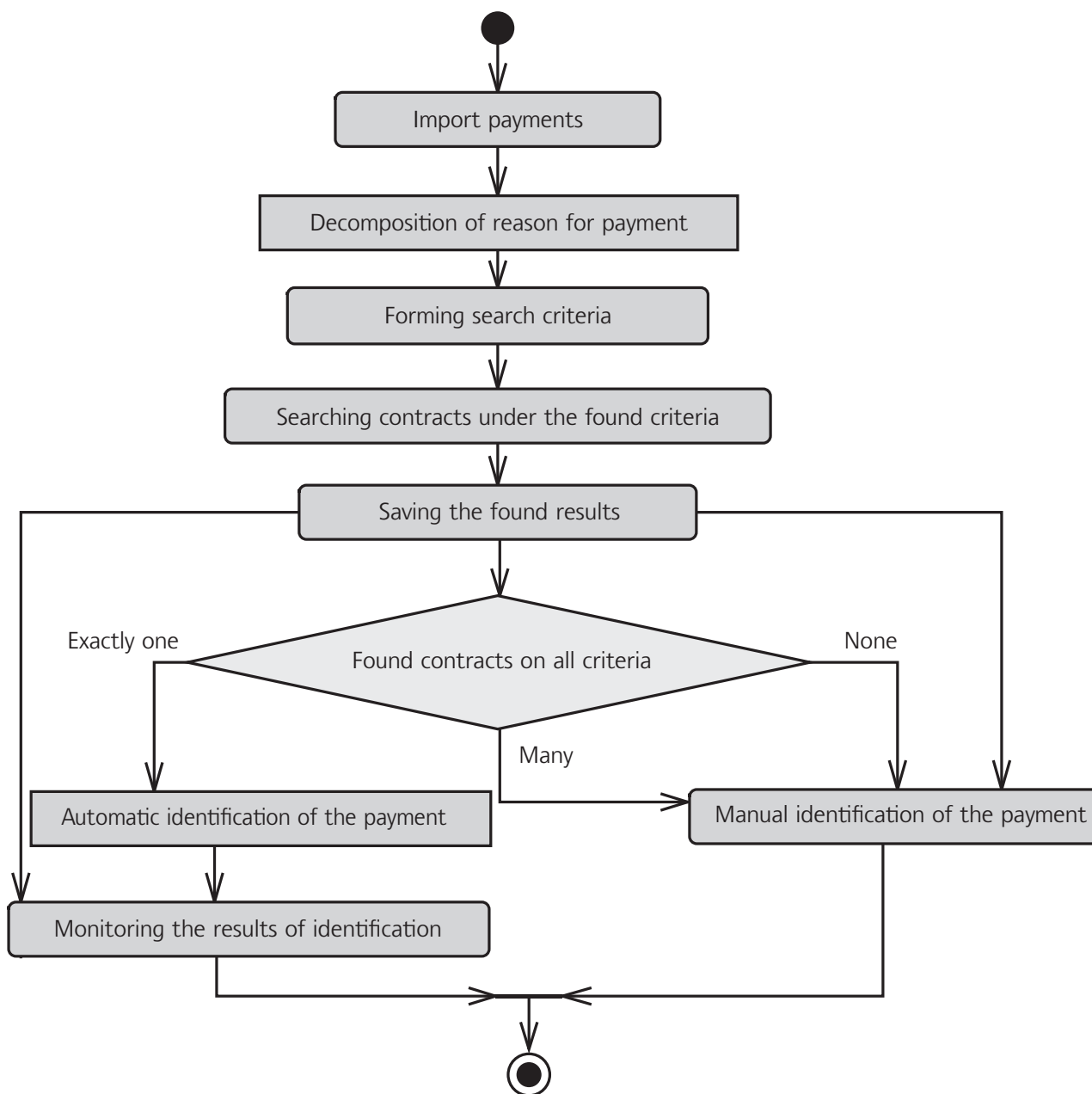

Figure 1. Business Process Model "A payment identification"

The essence of payments identification is in finding the specific contract in which the come in payment pays off the obligation. It is necessary the data available for the active contracts with customers to be compared with the filled in by the payer terms of payment.

The terms of payment is a free (unstructured) text which should be decomposed to separate potentially meaningful parts (words) and then to be analyzed and processed (identified). The purpose of decomposing and analysing the terms of payment is to separate those its elements containing client's identification features (two or three names, UCN, VAT number, phone number) or contract's identification features (number, the paid service or product type). Received payments can be definitely identified only by combining the results of analyzing the terms of payment and matching the amount of payment with the amount of past or upcoming contractual obligations. Using one of these indications of identification would lead to higher percentage of wrongly identified payments, because it is possible that:

• There are many customers with duplicate names, especially when only two names are filled in the terms of payments, not all three or four names.
• The amount of payment (respectively and the amount paid) to match for many contracts. Often customers pay a round amount which differs slightly from the actual debt and in this way the identification by amount paid is further hampered.
• There is a mistake in the number filled in.
• When using a ten digit contract number it is difficult to distinguish whether the customer has filled in the terms of payment a contract number or UCN, because it is possible the structure of the contract number to conform to a valid UCN and to prove that it belongs to a registered customer who operates another contract.

On the other hand, the more stringent identification criteria which avoid the above cases for wrong identification will reduce the number of successfully identify payments. The compromise between these two alternatives is determined by the corporate company policy (Figure 2).
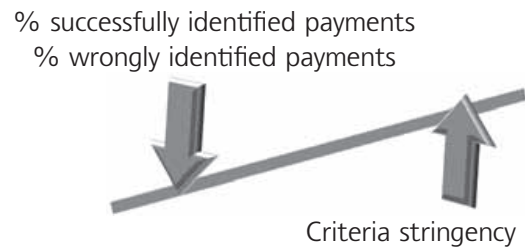
% successfully identified payments
% wrongly identified payments

Criteria stringency

Figure 2. Approaches to the criteria stringency for automatic payments identification

Regardless of the degree of compromise made between the identification performance (% of identified payments from the total number of processed payments) and its correctness (% of successfully identified payments against the total number of identified payments), in all cases a significant number of not identified payments remains after processing (between 5 and 20 percent), which must be identified manually.

The percentage of unidentified payments by the procedure is determined by two main factors:

• Efficiency of the implemented algorithm for analyzing plain text, including the text transformations avoiding common mistakes such as writing in Latin rather than Cyrillic (data in the main business application are stored in Cyrillic), case insensitive searching in text, etc.
• Correctness of filling in terms of payment incl. filling in all digits of the customer identifiers such as UCN or VAT number or ID of the contract, correct transcription of names, proper word separation in the terms of payment, etc.

Here arises the second, very important task that faces the automatic identification of payments – to preserve the results of the processed treatments in a convenient for the consumers form in order to make them usable for:

• Control over the automatic identification correctness;

• Subsequent manual identification of payments, which have remained unidentified or were incorrectly identified.

For useful and convenient usage of the results from the procedure for automatic identification execution must show:

• On which contract or customer attribute the searching was made;

• What was the searched value;

• Search results (found contract number);

• Further description of search results.

4. Ultimately, any payment would have to obtain a list of contract numbers that were found in searching different values in the role of different search criteria and clarification whether and why each result was used or ignored by the identification procedure. Such an approach would enable users to visually examine the terms of payment as they choose one of the following options:

• To link the payment with any of the contracts that the automatic identification procedure found in the search but did not ignore it for other reasons (for example because of non-coincidence of the amount of payment by the amount of duties or because of the fact that the found contract does not match the given conditions);

• As a last resort is doing manual search of contracts based on elements from the terms of payment, which does not repeat the searches that the treatment has already made, saving working time in this way.

As it will become clear from the statement in paragraph 5, storing the results of payments identification is essential to substantially reduce (in times) the labour-consumption of this process.

2. Methodological basis of the developed information technology

In Gartner' survey of the ten major information technologies to be developed in 2008 the accent is on "Metadata is the basis of the information infrastructure and is used by all the information systems. They are present in the databases for managing the configurations, in the registers of business services, in procedures for developing applications." [6] Metadata is data about data. Each item of metadata can describe a separate attribute or its content and also each item of metadata can describe multiple data incl. its contents. Metadata is used to facilitate understanding, using and managing of data, and may vary by its type and by context of use. It provides extra information about the data to the consumers, as this information can be as descriptive as well as algorithmic.

Depending on its purpose and method of use, it can be of one of two types – Structural (managing) and for searching support. Structural metadata is used for describing the structure of the computer systems incl. databases and their tables, columns, indexes, etc. The metadata for searching support is organized as a set of keywords of human language and is used to accelerate the content searching (especially in WEB) in a particular subject area. Relational database management systems have their own (embedded) mechanisms for metadata support (tables for description of the tables and the other database objects), also called catalogues. Custom metadata from "structural" type is used in the development process of the information technology for automatic identification of payments.

Another primary priority in the information technologies for 2008, which is necessitated by purely economic reasons, is the management

of business processes (BPM – Business Process Management). Business process management primarily is a result of the need of radically accelerating the business. Besides that it is a basis for building a service-oriented architecture (SOA) so necessary to the enterprises from the new generation in relation to their needs to respond flexibly to customer expectations and market changes.

The term "operational business process" affects the recurrent business process that are implemented in an organization in the context of its daily business, not business processes related to decision-making process carried out by senior management. Business processes management covers the activities carried out in an organization to manage and improve business processes.

Information systems for business process management make these activities faster and cheaper as they monitor their implementation, providing managers an opportunity to analyze and change processes based on monitoring data, not offhand. BPM is a management model that allows companies to manage the ongoing processes in them and assets belonging to them, and continuously improving them. A good process management allows the medium and large organizations to adapt their businesses to daily changes in the processes that are caused by competition, from changes in regulations or the challenges of the market, without needing constant support of their IT departments.

Initially, the focus of BPM was focused on automation of routine activities until recently its scope extends to the integration of processes governed by the people and where human intervention is carried out with sequentially or in parallel performed routine procedures.

The combination of these two guidelines is in the basis of the developed technology for automatic identification, such as:

• The proposed autonomous model of data (see Fig. 3) is integrated in the main database of the business information system using specially designed for this purpose structural metadata.
• The procedure for automatic identification of payments and registration of the results of its performance fully reflects the corresponding business process, and in addition it is configurable (again through appropriate metadata) to be able to be set as far as it is possible to the specificity and corporate policies of various enterprises. There are conditions for maximum comfort of human interference, which is required whenever necessary for decision making.

Table 1. Results from the execution of the procedure for automatic identification of payments

Column	Purpose
Number of payment	Primary data for the received payment which is used for comparison with results from the execution of the procedure for automatic identification.
Date of payment	
Amount of payment	
Terms of payment	
Date and time of the last identification	Gives guidance to the consumer when to start again the procedure for automatic identification to obtain better results because of significant changes in data (inserted or corrected data for customers or contracts).
Result from the identification	Shows if a payment is identified or not.
Results explanations	Shows the terms of payment on which the corresponding result is achieved in order the user to assess whether the result is correct or not.

3. Data model for storing the results of automatic identification of bank payments under contracts with customers

For easier controlling the correctness of the procedure for automatic identification

of payments, the consumers must be given a list of payments with the following example content (Table 1).

This list with the results from the automatic identification is created on the basis of data for payments (they are part of the business
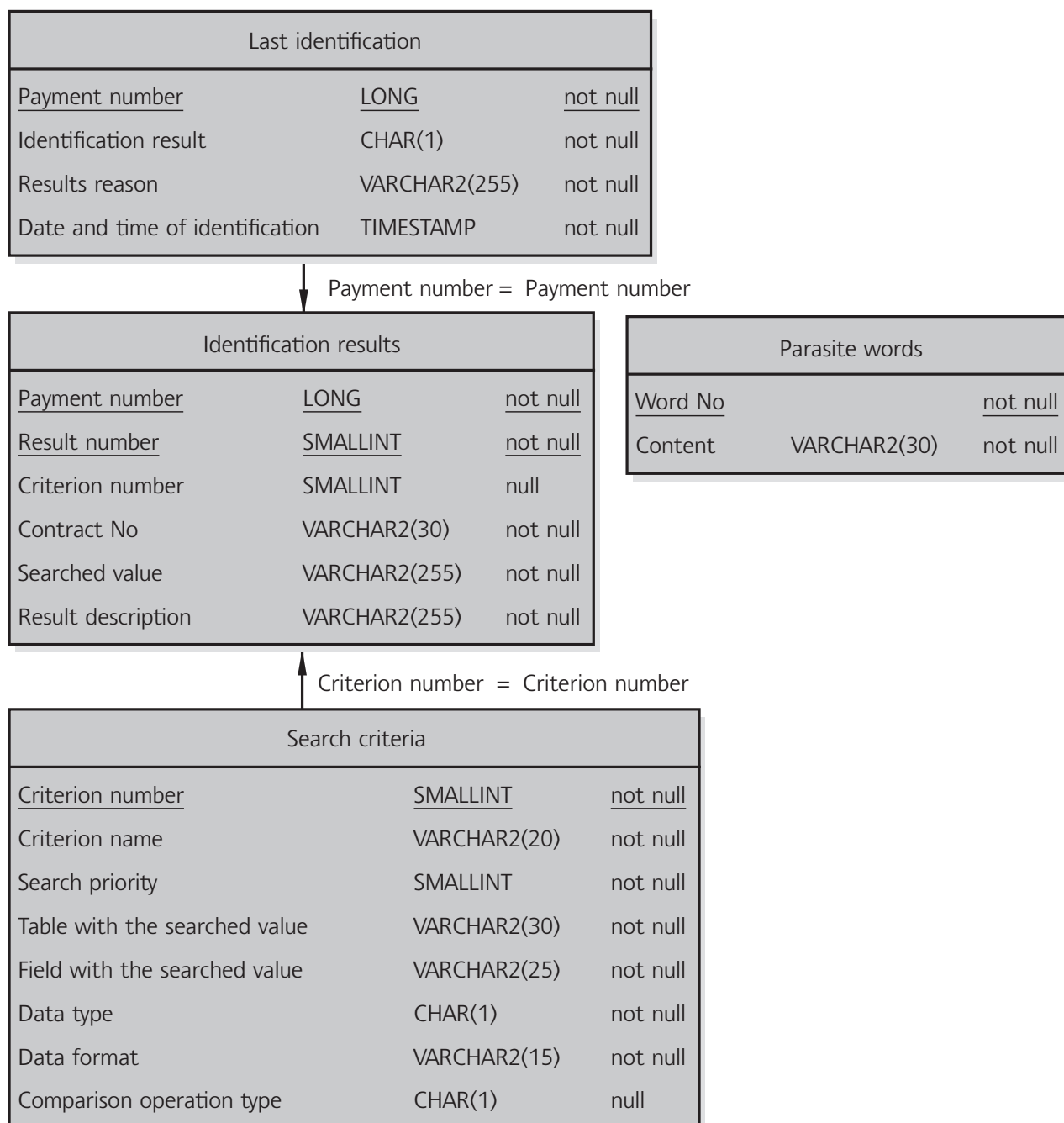
Last identification		
Payment number	LONG	not null
Identification result	CHAR(1)	not null
Results reason	VARCHAR2(255)	not null
Date and time of identification	TIMESTAMP	not null

Payment number = Payment number

Identification results		
Payment number	LONG	not null
Result number	SMALLINT	not null
Criterion number	SMALLINT	null
Contract No	VARCHAR2(30)	not null
Searched value	VARCHAR2(255)	not null
Result description	VARCHAR2(255)	not null

Parasite words		
Word No		not null
Content	VARCHAR2(30)	not null

Criterion number = Criterion number

Search criteria		
Criterion number	SMALLINT	not null
Criterion name	VARCHAR2(20)	not null
Search priority	SMALLINT	not null
Table with the searched value	VARCHAR2(30)	not null
Field with the searched value	VARCHAR2(25)	not null
Data type	CHAR(1)	not null
Data format	VARCHAR2(15)	not null
Comparison operation type	CHAR(1)	null

Figure 3. Physical data model for storing results from the automatic identification of bank payments

application and are not included in the model developed in Figure 3) and on the table "Last identification", in which for each payment is indicated whether it was identified or not and why (for example – a contract is found and it is with indicated in the terms of payment number, but the monthly fee it is not equal to the amount paid or is found a customer by 3 names, but he has several active contracts, the obligation under none of them match the amount of payment, so the payments stay unidentified or they are identified by criteria of the customer UCN and the amount of payment = amount of monthly payment, etc.)

As indicated in paragraph 2, for enabling the proposed data model to be integrated in the database of various business applications, it is necessary the search criteria for contracts using relevant elements of the terms of payment to be made as metadata. Following this approach in the data models development, a nomenclature is created, indicating each search criterion to which field of the database table must be set and what operation to perform for the comparison of the amount set aside for payment of ground with the value recorded in the DB.

The sequence of applying the criteria for searching contracts and clients can also be managed through this nomenclature, as to the more reliable criteria (such as contract number, UCN) a higher priority is assigned. Thus, if a payment is identified using a priority indicator, serious analysis on more unreliable data of two or three client names are avoided.

Searching results on each criterion is recorded in the table "Results from identification", whether the payment was identified or not. Later this data may be reviewed and to be determined exactly in the way the procedure have decomposed the terms of payment in parts, which from the received meaningful words with which criteria

for identification, and by what operation is compared and whether the number of contract to which the payment was linked was found as a result.

In order to accelerate the processing it is recommended some "parasitic" words to be excluded from the analysis of ground as meaningless if meeting them (such words as "ЕГН", "ДОГ" etc.) instead of searching on them to seek for customers or contracts.

4. Technology for processing automatic identification of bank payments under contracts with customers

Potentially meaningful (usable for payment identification) elements can be formed in the following algorithm:

1. The text is separated into words (sequence of symbols between two intervals or between any other two separator symbols except a dash).
2. Only words with 3 or more characters are separated- they could potentially contain a customer name.
3. So formed words are cleared of any eventually present parasitic words and are classified in 2 groups – textual, containing only letters and dashes and digital (only numbers from 0 to 9).
4. The search criteria of contracts in the database are retrieved and they are processed continuously according to the consistently given priority in the nomenclature (first most important) as:

•   The current criteria are applied consistently to each of the separated words in the terms of payment, which corresponds to its type and size or may be converted to the type and amount as set in the nomenclature format (for example, if Criterion 1 is a Number of

contract -15 digits, then all words of the terms for payment, which consist of 15 digits are searched using dynamic SQL in the indicated in the structural metadata field and table of the database.

• Data is inserted in the table "Results of identification". The inserted data indicates under what criteria and what value the search have been performed and also the contracts which have been found. For each contract result found, one record is added and if a contract was not found using the search criteria entered, the searching conditions are recorded and the field for the contract number remains blank. This is particularly significant for orientating users who will manually search contracts or customers by various criteria, avoiding repetitions of the treatments that have already been implemented by the procedure for automatic identification, but gave no result.

• For each entry found in the database it is verified that the forthcoming monthly contribution to the contract of the client corresponds to the amount of payment.

• If exactly one contract is found, and the amount of payment corresponds to the forthcoming instalment the payment is established as an identified one and the processing ends. Otherwise, the data searching continues using the next criterion of the nomenclature.

5. The processing ends after exhaustion of all the search criteria.

6. Searching results are summarized in order to form the basis of identification (just one contract by a definite criterion with a definite value found) or unidentified payment (not found any contract or more than one contracts which meet the criteria found).

In the description of that algorithm some technological details such as transformation of texts from Latin to Cyrillic and vice versa, ect. are not discussed.

5. Results from the application of the technology for automatic identification of bank payments under contracts with customers in a consumer funding company

The procedure for automatic identification of payments was implemented by the early 2006 in a company, which deals with target and consumer funding, and also with issuing credit cards. Currently the company has more than 580 000 clients, serving over 310 000 active contracts (some clients have no active contracts at the time).

Currently the procedure is configured to use strict criteria for identification as it is included a requirement for exact match of the amount paid with the amount of the obligation under the contract found. Furthermore, in order to improve the speed of the process, searching for customers and contracts by names (name, surname and family name or surname and first name) is proceeded only when no results from the more significant criteria have been received.

Priority of the search criteria is configured in the following way:

• Individual bank account number
• Contract number
• Customer UCN
• UCN of a related person (contact person)
• Name, surname and last name of the customer
• Name, surname and last name of a related person
• Name and last name of a client
• Name and last name of a related person

The following examples can serve for an illustration of the efficiency of the created technology in real terms:

1. Payment, in whose term a 10 digits number of contract cannot be separated (11 have been filled in as one of the digits is a duplicate) is identified in the 6th reliable criterion – client name (Figure 4). It should be noted that the procedure have identified the payment by three names of the client and by matching the amount of the payment with the obligations under the contract, while searching on two of the three names would have discovered 133 different contract, making the results unreliable and procedure does not use them.
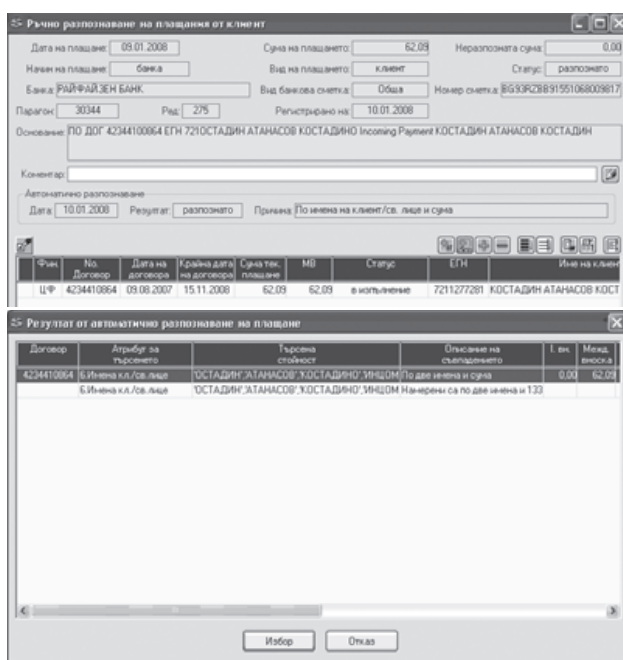
Figure 4. Form for a review of the saved results from a search for contracts with customers in the essential elements of the terms of payment (for the identified payment)

2. Payment in whose term there are two words with 10 characters is identified by both, as searching for value "4510306360" as a criterion "number of contract" has returned no result, while the same value sought by a customer UCN leads to finding a single contract. That contract is found on one more criterion – the 10 digit number of contract, making the results of identification largely accurate (Figure 5). In this case the procedure have not identified the

payment because it is configured to work only in exactly matching the amount of payment with the amount of the obligation under contract found.
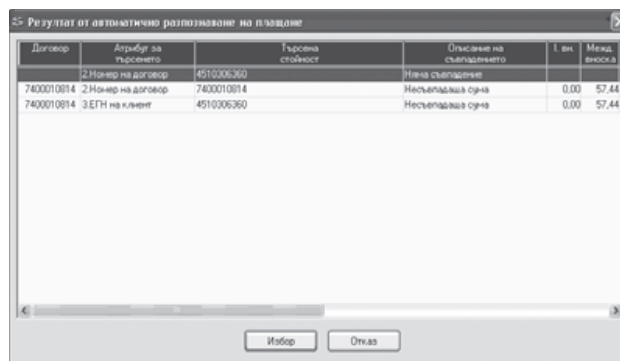

Figure 5. Form for a review of the saved results of searching contracts with customers by the essential elements of the term of payment
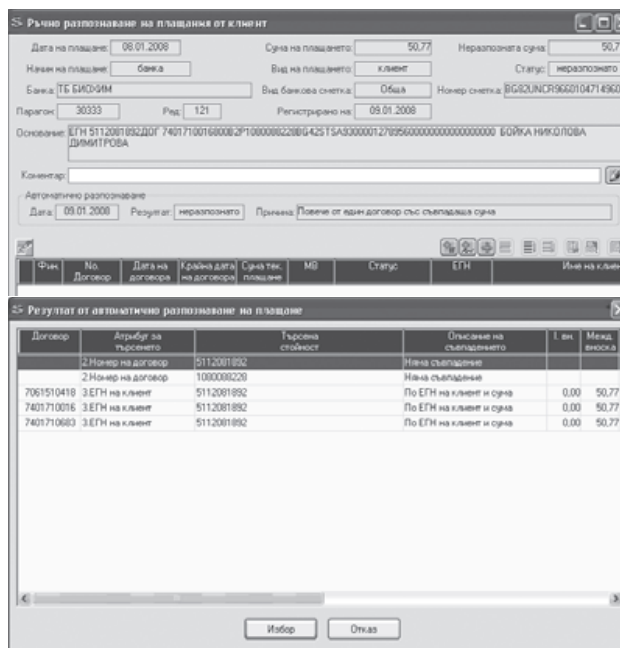

Figure 6. Identification of payment by customers with multiple active contracts

3. In cases when a customer has several active contract, but in the terms of payment only UCN is filled in correctly (as a contract number digits indicated are more than necessary), the procedure can not uniquely determine by which of the contracts the obligation will be paid off (Figure 6), as all three contracts are the same
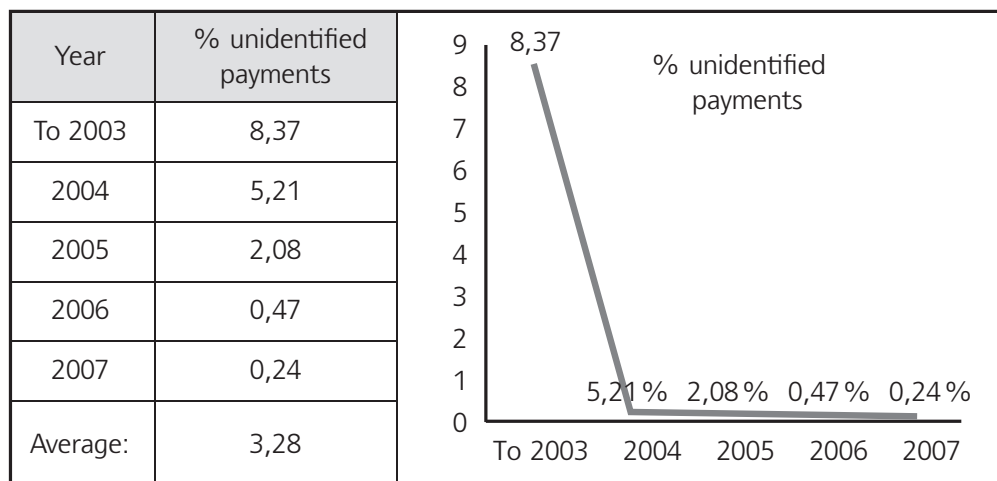
Year	% unidentified payments
To 2003	8,37
2004	5,21
2005	2,08
2006	0,47
2007	0,24
Average:	3,28

Figure 7. Results of implementation of the procedure for automatic identification of payments

customer's and the duties on them match the amount of payment. The users should choose one of the found out and listed contracts and to link the payment to it.

The results of implementation can be measured by two key indicators:

- % unidentified payments (Figure 7).
- Reducing the processing time of manual identification using the saved from the automatic identification results compared to the manual search of contracts in the database.

After its acceptance by the consumers the procedure was once executed on old data (from 2005 and earlier), which reduced the rate of unidentified payments from 12 percent before its execution to 5.21 percent after it, but unidentified payments processing for this period did not continue with manual detection because of lack of available human resources. More than 60 percent of employees in "Processing of payments were transferred to other departments because of the significant downsizing (more than 4 times) the time for manual detection payments using recorded in the database results compared to the automatical detection.

6. Conclusion

The presented information technology for automatic identification of bank payments under contracts with customers, which involves periodically occurrence and paying the obligations have already been used successfully for two years and clearly demonstrates its advantages in terms of efficiency algorithm and convenience for consumers.

This technology can be integrated to other business applications working in a broad class of companies and enterprises as it is realized on the basis of two of the leading trends in the development of information technologies, namely:

- Autonomous data models that can be built in as an extension of the database of any business information system, because it consists of structural metadata for a link to its tables and fields.
- Flexible procedure that can be configured on a set of criterion for identification of payments and the priority of each of them in order to adjust to the company policy and to the specific business process in various companies.

Literature

1. CIO, бр. 12, 2007, Платежните систе-ми на новото десетилетие – в посока към интеграция, http://cio.bg/?call = USE~home; &page = paper&n = 1641

2. CIO, бр. 12, 2007, Разплащателни услуги и възможности за интеграция в решенията за онлайн банкиране, http://www.cio.bg/?call = USE~home;&page = paper&n = 1646

3. CIO, бр. 12, 2007, АЖУР®–L5.5 в отра-съл енергетика – една година по-късно – http://www.cio.bg/?call = USE~home;&page = paper&n = 1455

4. CIO, бр. 4, 2007, Ефактура – възможност за спестяване на разходи -http://www.cio.bg /?call = USE~home;&page = resume&n = 1434

5. CIO, бр. 12, 2005, Информационни систе-ми за управление на бизнеса, http://www.cio. bg/?call = USE~home;&page = paper&n = 634

6. Gartner Identifies the Top 10 Strategic Technologies for 2008, http://www.gartner. com/it/page.jsp?id = 530109

7. Rascino, Mark, The Gartner Scenario: The current state and future direction of IT, Gartner Symposium/ITxpo Africa 2006, Cape town International Convention Centre, Cape Town, South Africa. ИА