

A METHOD FOR AUTOMATED PROCESSING OF INFORMATION

FIELD OF INVENTION

- 5 This invention relates to the processing of information extracted from field devices, in particular, the automation of the process.

BACKGROUND OF THE INVENTION

- 10 In systems for monitoring field devices in a Plant, messages generated and sent from the field devices are received by a message acquisition module in the plant server. The plant server includes a database for storing the field device information of the field devices. The message acquisition module then extracts the field device information from the incoming messages and then loads the necessary field device related information from the plant
15 database.

- In prior art message acquisition modules, the field device information is extracted from the messages with the help of a set of extraction patterns. The messages are in text format and the extraction pattern is simply a set of words which matches a part of the message in which
20 the field device information is included. A pool of patterns is constructed by the module developers on the basis of the knowledge of the existing message formats and hardcoded into the source code of the module.

- Fig 1 illustrates a prior art process flow 100 of the data extraction module. In Step 102, a
25 pattern is applied to an alarm message 104 to extract the field device information. In Step 106, the module determines whether the field device information has been extracted. If not, in Step 108, a next pattern from the pool 110 is loaded for the extraction. In Step 112, the module checks whether the pattern is the last one in the pool. If no, Steps 102, 106, 108 and 112 are repeated for this next pattern. The extraction process is ended when the field device
30 information has ended or no field device information is extracted when all the patterns have been applied to the alarm message. In the latter case, the module considers that no field device information is contained in the message.

- The extraction pattern differs for different field devices. Each time a new field device having a
35 new message format is added or the message format for an existing field device is modified, the developers have to update the source code to include the corresponding new or modified

extraction pattern. The module has to be modified and rebuilt to include new or modified patterns. This is not productive and inefficient. Accordingly, there is clearly a need for an improved method of performing the information extraction.

5

SUMMARY OF THE INVENTION

This invention is a method and system for automating the information extraction to reduce the manual efforts required to maintain the message acquisition module.

10

A first embodiment of the invention, there is a method for extracting field device information from a field device message in a system having a plant database of field device information. The method comprises the steps of tokenizing the field device messages into a pool of tokens, constructing a pattern with all the tokens in the pool, applying the constructed pattern to extract field device information from the message, and verifying extracted field device information.

15

A second embodiment, there is a method for extracting field device information from a field device message in a system having a plant database of field device information and a pool of patterns formed of tokens. The method comprises the steps of tokenizing the field device messages into a pool of tokens, constructing a pattern with all the tokens in the pool, applying the constructed pattern to extract a first field device information from the message, verifying the first extracted field device information. When the first extracted field device information is correct, the method further includes unifying the constructed pattern with new pattern, unifying the constructed pattern with new pattern, constructing a unified pattern using the new pattern and one old pattern from the pool, applying the constructed unified pattern to extract a second field device information from the message and verifying the second field device information.

20

25

30 Preferably, the tokenizing step is done according to Treebank Tokenization.

Preferably, the step for verifying the extracted field device information further includes loading the field device information from the plant database, to determine if the extracted field device information is correct.

35

Preferably, in the first embodiment, in the step for verifying the extracted field device information, if the field device information is verified to be not correct, the pattern is reconstructed by stripping the unnecessary tokens from the pool.

- 5 Preferably, in the second embodiment, in the step for verifying the first extracted field device information, if the field device information is verified to be not correct, the pattern is reconstructed by stripping the unnecessary tokens from the pool.

10 Preferably, the unifying step is done by unifying the tokens that are common to both the new and old pattern.

Still preferably, the unifying step is done by setting a threshold for the number of common tokens being used.

- 15 Preferably, in the second embodiment, in the verification step for the second extracted field device information, when the field device information is verified to be not correct, the pattern is reconstructed by unifying the new pattern with a different old pattern from the pool.

20 Preferably, in the second embodiment, in the step for verifying the second extracted field device information, when the field device information is verified to be correct, the pattern is reconstructed by unifying the unified pattern with a different old pattern from the pool for reducing the tokens used in the construction.

25 In another embodiment, the invention is a system for extracting field device information from a field device message. The system comprises a database containing field device information, a database of pattern which are formed of tokens, a tokenization module for tokenizing the message into a pool of tokens, a temporary storage for the pool of tokens, a bootstrapping module for constructing a pattern from the pool of tokens and applying the constructed pattern to extract field device information. Finally, a verification module for
30 verifying the extracted field device information.

BRIEF DESCRIPTION OF THE DRAWINGS

35 Figure 1 is a process diagram illustrating the prior art.

Figure 2 is a process diagram illustrating the first part of the preferred embodiment.

Figure 3 is a process diagram illustrating the second part of the preferred embodiment.

5 Figure 4 is a diagram of the system for the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

10 Figure 2 is an illustration of one preferred embodiment of the invention, which is, a method 200 for extracting field device information from a field device message. The field device information is the field device path information or maintenance information from alarm messages such as a problem, cause. Step 202 applies a Tokenization Model to a field device message 204. In the Tokenization Model, a tokenizer 206 breaks up the message into
15 an output pool of tokens 208. This can be done differently at various levels by the different algorithms available. A text can be broken up into paragraphs, sentences, words, syllables or phonemes.

In an example for extracting field device path information from the field device message
20 "fielddevice\$area@location.plant", the tokenization model chosen by the Tokenization Model is a word, the pool of tokens include "fielddeivce", "area", "location" and "plant. If special symbols such as "@", "\$" are defined in the model, the pool of tokens include the special symbols which are useful for constructing a pattern in a later step.

25 Preferably, a Treebank Tokenization is applied to the model. This tokenization is free of charge, has an acceptable level of accuracy in results, and is easy to implement.

In Step 210, the pool of tokens is input to a Bootstrapping Model. In the Bootstrapping Model, a bootstrapping module 212 constructs a pattern and applies the constructed pattern to the
30 message 204 to extract the field device information 214.

In Step 216, the extracted field device information 214 is applied to a Verification Model. In the Verification Model, a verification module 218 loads the corresponding field device information which is using it from the plant database. If the corresponding field device
35 information loading fails, the extracted field device path information is not correct. In Step 220, an error is returned during the verification and a token is removed from the pool. The

Steps 210 and 216 are repeated until the correct field device information is found. If the extracted field device path information is correct, in Step 222, the pattern learnt is stored in a database of patterns. Step 224 determines whether a correct field device path information has been found. If not, the Steps 210 and 216 are repeated.

5

With the above automation, the efficiency of extracting information from field devices and the productivity of the module developers are improved.

10 Still preferably, the pattern learnt is generalised with the existing patterns in the database of patterns. It is likely that a newly learnt pattern shares some common tokens with some existing patterns. By combining the common tokens of two patterns, a more generalised new pattern is created. The new pattern is able to match messages which are extracted by two original patterns previously. Since the new pattern provides a wider coverage on the extracting of the field device path information, it is advantageous to replace the original two
15 patterns with the new pattern in the database of patterns. This will further improve the efficiency of extracting field device path information and optimize the utilization of the database of patterns.

20 The generalization model 300 is illustrated in Fig 3. In Step 302, a newly learnt pattern 304 is input into a Unification Model. In the Unification Model, Step 306 unifies one existing pattern in the database 308 with the new pattern. In the unification step, at least one threshold is set to ensure that the unified pattern is not so generalized that the accuracy of the field device path information is comprised. In an example threshold, one pattern has to consist of at least five tokens. In the generalization process, the unification process stops when the new pattern
25 reaches five tokens length.

In Step 310, the database of patterns is input to a second Bootstrapping Model. In this second Bootstrapping Model, a second bootstrapping module 312 constructs a unified pattern and applies the constructed unified pattern to the message 204 to extract the field
30 device path information 314.

In Step 316, the extracted field device path information 314 is applied to a Verification Model. In the Verification Model, a verification module 318 loads field device information from the plant database. If the extracted field device path information is not correct, in Step 320, an
35 error is returned during the verification and the unified pattern is not stored. Steps 302, 310 and 316 are repeated for the newly learnt pattern 304 with the next existing pattern in the

database. If the extracted field device path information is correct, in Step 322, the unified pattern is unified with the next existing pattern in the database. The Steps 302, 310 and 316 are repeated until Step 324 determines that all the existing patterns in the database have been processed. The unified patterns are updated to the database of patterns 326.

5

In yet another embodiment, the invention is a system 400 illustrated in Fig 4. The system includes database 402 containing field device information, database of patterns 404 which are formed of tokens. Messages from field devices 406 are input into the Tokenization module 408 to generate a pool of tokens. The generated pool of tokens is stored in a temporary storage 410. The pool of tokens is input into a Bootstrapping Module 412 for constructing a pattern from the pool of tokens and applying the constructed pattern to extract the field device path information. The corresponding field device path information from the field device path database is loaded into the Verification Module 414 for verifying whether the extracted field device path is correct and storing the correct pattern into the database of patterns 404.

10
15

It is to be understood that the foregoing description of several preferred embodiments is intended to be purely illustrative of the principles of the invention, rather than exhaustive thereof, and that changes and variations will be apparent to those skilled in the art, and that the present invention is not intended to be limited other than expressly set forth in the following claims.

20

We claim:

1. A method for extracting field device information from a field device message in a system having a plant database of field device information, comprising the steps of:
 - (a) tokenizing the field device messages into a pool of tokens;
 - (b) constructing a pattern with all the tokens in the pool;
 - (c) applying the constructed pattern to extract field device information from the field device message; and
 - (d) verifying the extracted field device information, when the extracted device information is verified to be not correct, the pattern is reconstructed by stripping the unnecessary tokens from the pool.
2. A method for extracting field device information from a field device message in a system having a plant database of field device information and a pool of patterns formed of tokens, comprising the steps of :
 - (a) tokenizing the field device messages into a pool of tokens;
 - (b) constructing a first set of patterns with all the tokens in the pool;
 - (c) applying the constructed first set of patterns to extract a first field device information from the message;
 - (d) verifying the first extracted field device information when the first extracted device information is verified to be not correct, the pattern is reconstructed by stripping the unnecessary tokens from the pool;
 - (e) wherein the first extracted field device information is correct, the method further includes;
 - (f) combining one of the first set of patterns and a newly inputted pattern to construct an unified pattern;
 - (g) applying the constructed unified pattern to extract a second field device information from the message; and
 - (h) verifying the second field device information.
3. A method as claimed in claim 1 or 2, wherein tokenizing is done according to Treebank Tokenization.

4. A method as claimed in claim 1 or 2, wherein the step for verifying the extracted field device information further includes loading the field device information from the plant database to determine if the extracted field device information is correct.
5. A method as claimed in claim 2, wherein the unifying step is done by unifying the tokens that are common to both the one of the first set of patterns and the newly inputted pattern.
6. A method as claimed in claim 2, wherein the combining step is done by setting a threshold for the number of common tokens being used.
7. A method as claimed in claim 2, wherein the step for verifying the second extracted field device information, the second extracted field device information is verified to be not correct, the unified pattern is constructed by combining a different one of the first set of patterns and the newly inputted pattern.
8. A method as claimed in claim 2, wherein the step for verifying the second extracted field device information, the second extracted field device information is verified to be correct, the unified pattern is constructed by combining a different one of the first set of patterns and the newly inputted pattern for reducing the tokens used in the construction.
9. A system for extracting field device information from a field device message comprising:
 - a) a database containing field device information;
 - b) a database of pattern formed of tokens;
 - c) a tokenization module for tokenizing the message into a pool of tokens;
 - d) a temporary storage for the pool of tokens;
 - e) a bootstrapping module for constructing a pattern from the pool of tokens and applying the constructed pattern to extract field device information; and
 - f) a verification module for verifying the extracted field device information and stripping unnecessary tokens from the pool when the extracted device information is verified to be wrong.