

Talking technology

R&D director Jörkki Hyvönen answers questions about mCorrection Intelligent Address Lookup, and discusses how the software improves read rate and lowers costs



Jörkki Hyvönen

Jörkki Hyvönen is a R&D director and partner at Syslore. Hyvönen has an extensive background in academic and scientific research, especially in engineering physics. He received his MSc in engineering physics from the Helsinki University of Technology (HUT). His scientific career includes research at the HUT Laboratory of Biomedical Engineering and Low Temperature Laboratory. Hyvönen is currently completing his PhD research at the HUT Laboratory of Computational Engineering.

What is mCorrection Intelligent Address Lookup?

mCorrection Intelligent Address Lookup is a software solution that can act as a complementary and secondary lookup solution to upgrade the performance of existing sorting systems and OCR engines, or it can be integrated to replace the OCR system's built-in address lookup function to boost its recognition efficiency and accuracy.

How does mCorrection Intelligent Address Lookup differ from other address

recognition and lookup systems?

mCorrection Intelligent Address Lookup software uses a different approach to address recognition compared with conventional OCR address lookup and recognition systems. With other address lookup systems, the delivery point is the most error prone of all sorting levels. In other words, the letters automatically sorted all the way to the final delivery point are also the ones that are most likely to be delivered to a wrong address.

With mCorrection Intelligent Address Lookup, this is reversed: the letters sorted to the delivery point are also most likely to be delivered to the right address. This is only logical, because in mCorrection, exact and inexact data is used in the matching process. The delivery point is the level containing the most data: the street name, the house number, the city name, and possibly even the name of the recipient. When more data is available, cross-referencing for accuracy is easier.

With mCorrection Intelligent Address Lookup, life extension and upgrade projects for the existing sorting and OCR infrastructure can be made with significantly lower costs and in a shorter time-frame compared with OCR investments. Achieving recognition improvements of literally tens of percentage units by deploying secondary recognition software running on standard hardware is definitely a point of differentiation.

Tell me about the software technology behind the address lookup solution.

mCorrection is the first database index developed for fuzzy string searches, which is applicable for the whole edit distance family of error metrics. I believe that our team has truly made a breakthrough in

solving scalability problems and combined the solutions with the ability to maintain a tight accuracy level in the matching process. mCorrection fuzzy search time achieves sub-linear scalability even with very loose match quality limits, outperforming existing Levenshtein and similar dynamic programming solutions by orders of magnitude.

How was mCorrection matching technology originally invented?

mCorrection fuzzy matching technology was originally invented to solve usability and address matching issues of a certain route optimization software. The problem was that there were a lot of unrecognized (and uncharged) search requests because of misspelled or abbreviated street and city names.

The Syslore R&D team adapted our experience from bioinformatics to solve the problem, and invented mCorrection, which virtually eliminated address matching problems because inexact and ambiguous service requests were now automatically recognized. After the technology had been further developed and commercialized, its benefits in detailed multicolumn matching were applied in postal address matching with spectacular results.

Tell us a little bit more about mCorrection address lookup solution. What are its advantages in Optical Character Recognition (OCR) based address matching and in the mail sorting process?

In mail sorting, mCorrection provides three main benefits to the automated sorting process. First, it improves the current address lookup solution's read-rate, especially to the delivery point level, leading to reduced manual mail sorting.

"MATCHING OF EXACT AND INEXACT DATA MAKES CROSS-REFERENCING EASIER"

Secondly, mCorrection can, in real time, reduce the number of recognition errors the current OCR solution makes. Hence, it efficiently cuts down the number of misdirected mail items. The third benefit is that mCorrection can find matches in real time for the addresses the current solution has rejected, thus reducing manual video coding.

Automatic mail sorting and address matching can be different and of variable quality in different countries. Does mCorrection need a lot of tailoring for different languages, address databases, or OCR and sorting infrastructures?

From the beginning of the development, we were prepared for different address and language/character requirements. Finland is a bilingual country, with Finnish as the majority language, and a small minority speaking Swedish. This poses some challenges to our address system, because the majority of street names in the Finnish postal system have

two official names (Finnish and Swedish) with several variations and abbreviations widely used. This is perhaps one of the main advantages in our technology compared to others; our fuzzy matching algorithms solve these problems intelligently by allowing efficient use of abbreviations and alias-words, and some other specific techniques and algorithms, in the recognition process.

We also made the decision at the beginning to develop the software as a pure Java solution due to its UNICODE character set support, and the ability to compare any mix of characters. This has proved to be a help in keeping testing and implementation projects straightforward. Also, integration to additional sorting centers and units is an easy task.

You mentioned about the scalability of your technology. What are mCorrection address lookup's response times in the mail sorting environment? Do huge address databases and data amounts

affect the performance of your solution?

mCorrection Intelligent Address Lookup response times depend on the quality of the OCR system's output. In practice, response times vary between 15–200 milliseconds, and are fully adjustable. Therefore, all mCorrection Intelligent Address Lookup matches can be executed online without delays or changes in the existing mail sorting process. mCorrection has been tested with database sizes of over one billion rows and it has maintained its outstanding performance. We look forward to be challenged with large data sets and customers with scalability and accuracy problems in address matching. ■

FIND OUT MORE

Syslore Ltd
tel: +358 9 5627 5450
email: info@syslore.com
web: www.syslore.com