A New Topological Method for Examining Historical Inscriptions

Loránd Lehel Tóth, Budapest University of Technology and Economics, Budapest, Hungary
Gábor Hosszú, Budapest University of Technology and Economics, Budapest, Hungary

ABSTRACT

The article presents a new method developed to increase the efficiency of the identification algorithm for historical inscriptions of unknown origin. The authors extracted topological properties of the symbols containing different script relics, and analyzed them by using statistical tools. The considered topological properties are circular loop, oblique lines, vertical section, and crossing, among others. The article describes the use of the number of a three-circle grapheme intersection point vectors to identify unknown symbols. The number of intersections of the three circles and the examined symbol is stored in a feature vector. By supplementing the feature vectors with the circle vectors, the authors succeeded in improving the efficiency of the algorithm designed to decipher hard to read historical inscriptions.

KEYWORDS

Computational Paleography, Engineering in Paleography, Grapheme Analysis, Human-Computer Interface

INTRODUCTION

The humanities-based paleography is a science that deals with the reading of ancient writing systems (scripts). Differently, computational paleography, in other words engineering in paleography, as a branch of applied computer science, deals with the spatial analysis of various glyphs, models the evolution of different scripts (Hosszú & Kovács, 2016), and provides a support for deciphering ancient inscriptions (Tóth et al., 2015), among others. Computational paleography extends the engineering modeling methods to any data of the written cultural heritage.

In this article the basic written units of the scripts are called graphemes, in other word characters. A grapheme is the smallest semantically (Sukkarieh et al., 2012) or phonetically distinguishing element in a writing system. The grapheme can be of different types: letters, ligatures, pictograms, ideographs, numbers, logograms, punctuation marks, etc. The graphemes can have various properties. They have a transliteration value in angle brackets, one or more shapes called glyphs, one or more sound values between slashes, and terms of usage. Among the glyphs there is one that was used to be treated with highest priority; it is usually called a typical glyph. A normalized (“ideal”) glyph is a designed typical glyph of a grapheme in accordance to the most significant visual properties of a
grapheme. The most significant visual properties of a grapheme are called visual identity. The visual identity, the topology, the phonetic meaning and the semantic usage of a grapheme constitute the components of a layered grapheme model developed by Pardede et al. (2016).

The productions of using a writing system are the so-called inscriptions. Inscriptions—indeed, the applied writing technology (carving a wall, writing onto paper, etc.)—is generally composed of symbols, which are the smallest individual units of the inscriptions from a visual perspective. Consequently, a symbol is the materialization of a particular glyph of a grapheme, and the glyph of a grapheme is the abstraction of a symbol (Hosszú, 2014). The symbols and glyphs are called together shapes.

The task of examining ancient inscriptions is largely mapping symbols of the inscription to be deciphered to the grapheme set of a certain script. This problem differs from the goal of the well-known optical character recognition (OCR). Namely, while in case of OCR it can be assumed that normalized glyph of the written symbols is well-known and visual information found in the inscription should be assimilated to some well-known grapheme, the visual information on the other hand found in the inscription during examining historical inscriptions has to be assigned to a grapheme in such a way, that the typical glyph of that grapheme in a certain age is not known either. In the performed examinations the visual information of the glyphs is described by providing set of topological parameter values suitably chosen by human intervention and the resulting set of parameters is used as a series of input data for identification procedure.

The reading and interpretation of the inscriptions made in ancient times causes several difficulties for researchers. Except the deterioration of the writing base material (wood, stone, brick, paper, etc.) the reason for this is that the glyphs used in the process of writing have changed over the time as well as various inscriptions were created by people with a different knowledge and different handwritings. We have been developing a complex algorithm for deciphering inscriptions without reading (Tóth et al., 2015, 2016). This deciphering algorithm basically compares the topology of each symbol of the inscription to be deciphered to the topology of each known glyph of the graphemes of a database. The main difficulty is to find the most similar known glyphs to each symbol of the inscription to be deciphered. The inherent variance of the shapes of the symbols of an inscription limits the reliability of finding the appropriate glyphs. The symbols of an inscription are presumably the realizations of graphemes; therefore, the visual identities of the glyphs of the graphemes in a script used for making this inscription and the visual identities of the symbols could be identical; however, the actual topologies of the glyphs of the appropriate graphemes and the symbols in the inscription could be largely different. In order to overcome this difficulty, for each symbol the set of the best fitted glyphs is selected instead of choosing only one glyph. Using the combinations of the selected glyphs, different deciphering attempts are generated, and these deciphering attempts are evaluated based on a dictionary and grammar rules of the presumed language of the inscription. Finally, the reliability of the best deciphering attempts are calculated as the output of our deciphering algorithm. In this procedure, the comparison of the symbols of the inscription to the known glyphs is a crucial step. In order to improve the reliability of the estimated topological parameters of the symbols in the inscription, a new preprocessing method called Circle Method has been developed as an addition of our complex deciphering algorithm.

The Circle Method as a preprocessing step is not script-specific; therefore, it can be applied for any writing system. However, the Circle Method was tested only for deciphering the medieval relics of the Székely-Hungarian Rovash (SHR). The SHR was mainly used by the Székelys (Hungarian ethnic group in Transylvania, in Romania) as an unofficial writing system up to the 19th century (Hosszú, 2013, 2017). The SHR indicates sounds according to the alphabetic principles: each sound generally corresponds to one grapheme. In most cases, the line written in SHR proceeds from right to left, but less frequently it is possible to see ancient texts where it can be proceeded from left to right. In the latter case, the glyphs reflect horizontally. There was no distinction made between lowercase and uppercase letters.
Related Content

Gg
[www.igi-global.com/chapter/gg/76416?camid=4v1a](www.igi-global.com/chapter/gg/76416?camid=4v1a)

Measuring Information Success at the Individual Level in Cross-Cultural Environments
[www.igi-global.com/article/measuring-information-success-individual-level/51028?camid=4v1a](www.igi-global.com/article/measuring-information-success-individual-level/51028?camid=4v1a)
Intrusion Detection Based on P2P Software
[www.igi-global.com/chapter/intrusion-detection-based-p2p-software/13891?camid=4v1a](www.igi-global.com/chapter/intrusion-detection-based-p2p-software/13891?camid=4v1a)

Neural Networks for Retail Sales Forecasting
[www.igi-global.com/chapter/neural-networks-retail-sales-forecasting/13986?camid=4v1a](www.igi-global.com/chapter/neural-networks-retail-sales-forecasting/13986?camid=4v1a)