

Üveges István

## **Közérthetőség és automatizáció**

### **- kísérletek a jog, természetesnyelv-feldolgozás és informatika határán**

#### *Kivonat*

A jogi szövegek természetükből adódóan olyan magatartási normákat határoznak meg, amelyek betartása mindenkire nézve kötelező, és amelyek ismertét és megértését a jogalkotó minden állampolgár részéről feltételezi és elvárja. Ennek ellenére az ilyen szövegek megalkotása során legtöbbször csak a szaknyelvet ismerők szempontjai érvényesülnek, a laikusok nézőpontja nem. A problémára megoldást jelenthet, ha a közérthető fogalmazás automatizációs lehetőségeit vizsgáljuk.

Egy olyan szoftver, amely egyfajta helyesírás ellenőrzőként segít a fogalmazónak könnyebben érthetővé tenni a szöveget, jelentős emberi munkaerő megtakarítását teszi lehetővé. Az értekezés a jogi és nyelvészeti szakirodalomra támaszkodva keresi a választ arra a kérdésre, hogy mely nyelvi, nyelvhasználati jegyek megváltoztatása segítheti a jogi szövegek közérthetővé tételét. Ennek során a kutatás nagyban támaszkodik a Plain Language Movement által meghatározott alapelvekre és javaslatokra. Egyszersmind felügyelt-, és felügyelet nélküli gépi tanulási megoldásokkal vizsgálja a közérthetőséget, mint klasszifikációs problémát.

A kutatás eredményképpen elkészített szoftver egyfajta prototípusként hivatott szolgálni, amely segítheti a témával foglalkozók munkáját, ha saját megoldás elkészítése mellett döntenek. Egyúttal a megfelelő környezetben alkalmazva segítséget hivatott nyújtani a közérthető nyelvhasználatot fontosnak tekintők számára, gyorsítva főként a funkcionális jogi szövegek (pl. hivatali tájékoztatók) közérthető átiratainak elkészítését.

Üveges István

## **Comprehensibility and automation**

**- experiments at the interface of law, natural language processing and informatics**

### *Abstract*

Legal texts, by their very nature, lay down standards of behavior that are binding on all, and which the legislator presupposes and expects all citizens to know and understand. Nevertheless, in most cases, such texts are drafted from the point of view of those who know the legal language and not from the layman's perspective. A solution to this problem could be found by exploring the possibilities of automating the drafting of Plain Language.

Software that acts as a kind of spell-checker, helping the author to make the text easier to understand, could save considerable human labor. The thesis draws on the legal and linguistic literature to answer the question of which linguistic features can be changed to make legal texts easier to understand. In doing so, the research draws heavily on the principles and proposals set out by the Plain Language Movement. At the same time, it investigates comprehensibility as a classification problem using both supervised and unsupervised Machine Learning solutions.

The software produced as a result of this research is intended to serve as a kind of prototype, which can help those working in the field if they decide to develop their own solutions. At the same time, if used in the right environment, it should help those who consider the use of Plain Language important, by speeding up the production of plain language transcripts of functional legal texts (e.g. administrative information).