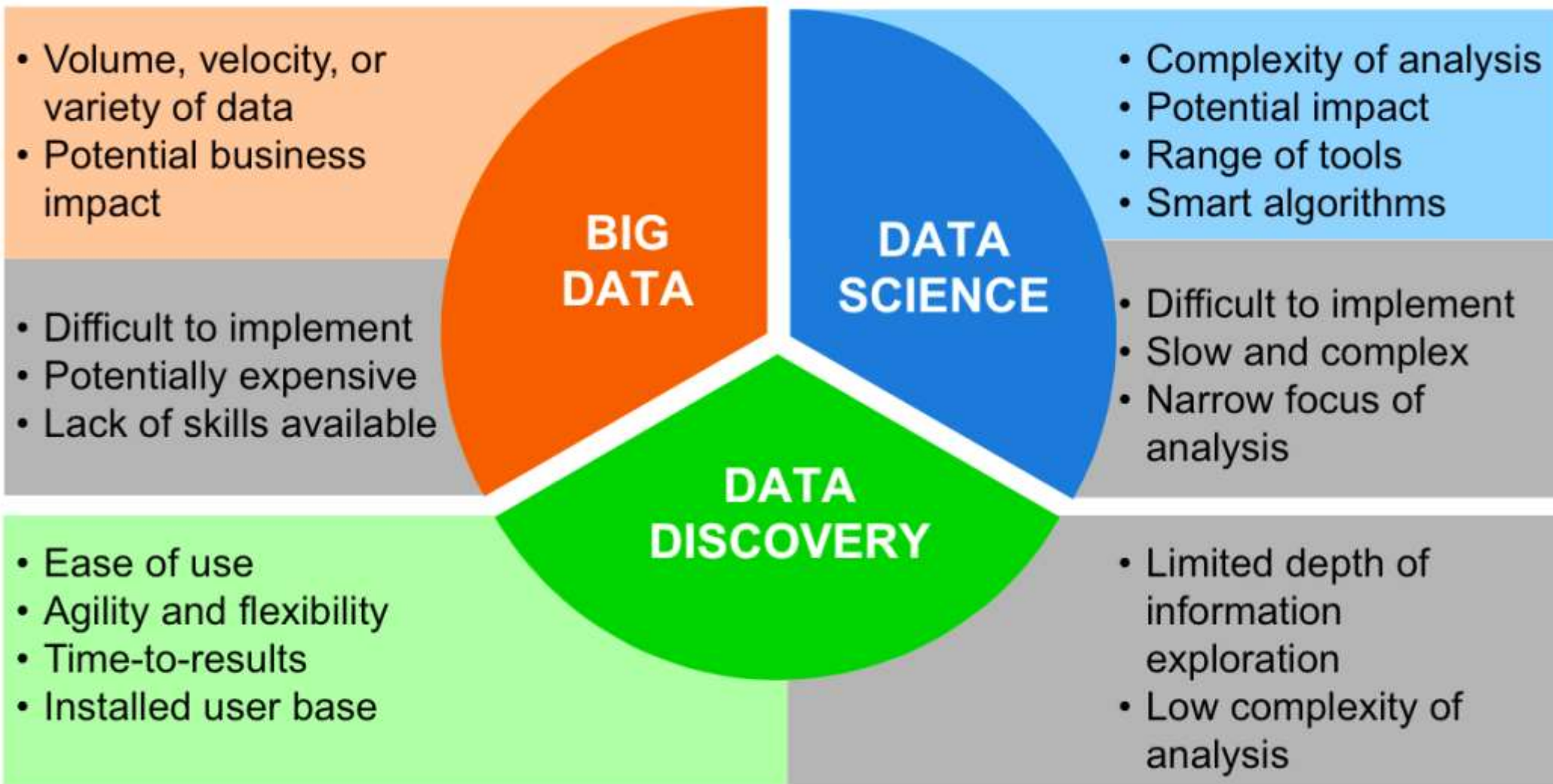


Személyiségvonások kinyerése strukturálatlan szövegekből:
a Personogram projekt

Big Data, mint új tudományos paradigma



Big Data, Data Science, Data Analysis



Adatforrások méret szerint

- **Üzleti adatok**
- **Humán adatforrások**
- **Gépi eredetű adatok**



- Mi a Big Data?
- Leginkább interpretációs keret.



A Big Data eddig leginkább a társadalomtudományok területén mutatott fel új eredményeket

- Leíró jelleg, szemben a kauzalitás alapelveivel.
- Nem egységes alapokon nyugszik.
- Az eredmények könnyen átültethetők a gyakorlatba.
- Nincsenek alapaxiómák, amik az adat vezéreltség ellen hatnának.

Példák: sör – pelenka, zseblámpa – cukorka

ajánlórendszerek, szociális média algoritmusai (matching)

Adatok a szociális médiából

- Facebook
- Twitter
- LinkedIn
- Blogbejegyzések,
- Kommentek (Disqus)
- Személyes dokumentumok
- Internetes keresések
- Webáruházak
- Médiafogyasztási szokások (ajánló rendszerek)
- Instagram
- Youtube
- Pinterest

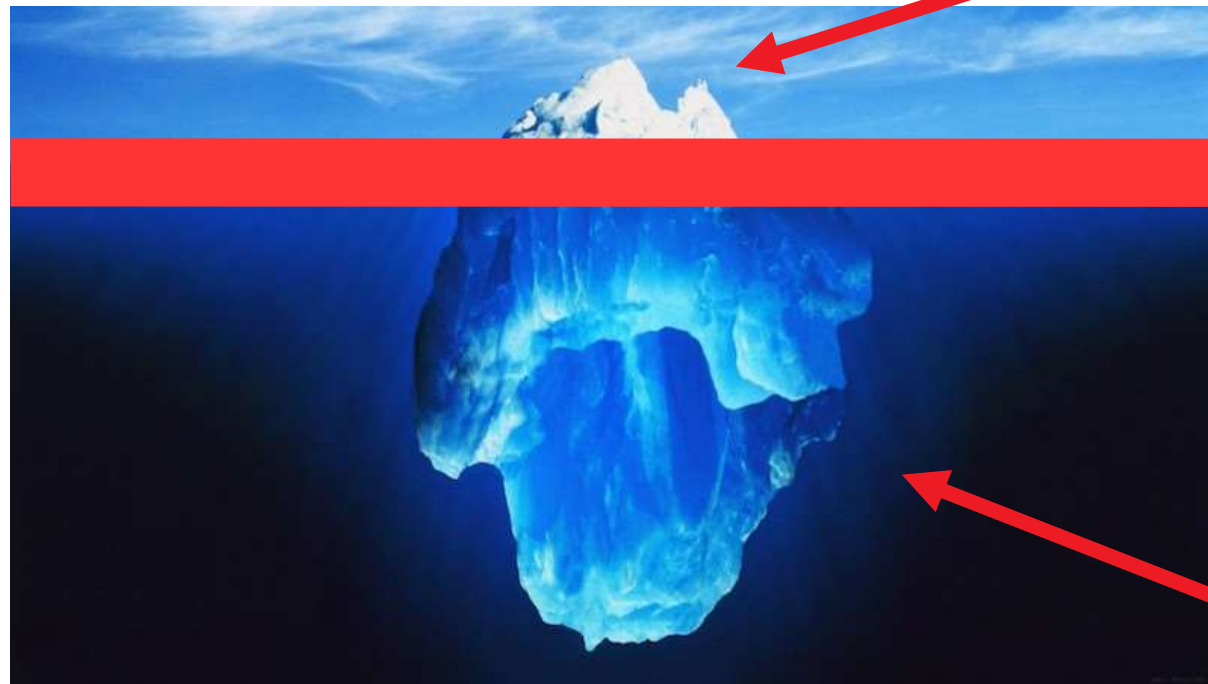


A viselkedésünk digitális leképeződése az áru.



Humán személyiség, vagy perszonifikált entitás többé-kevésbé invariáns, emberi viselkedéshez köthető (hasonlítható) tulajdonságainak összessége.

Jelen esetben a személyiség fogalma a viselkedésünk digitális leképeződése.



Small data eszközök
és módszerek

Teljes személyiség

Big data eszközök
és módszerek

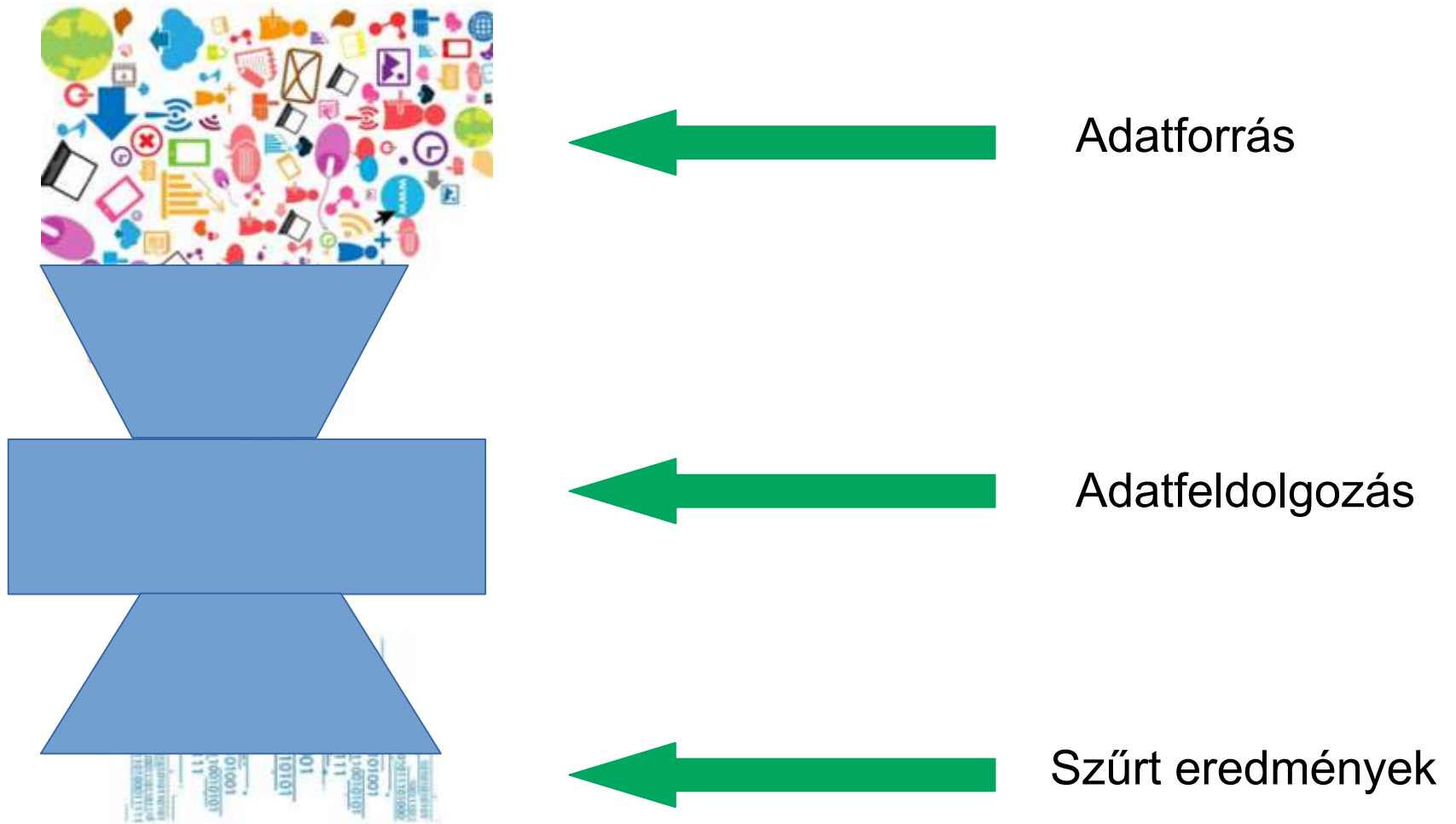
Ismeretlen reprezentáció, ismeretlen szabályokkal



Az általuk fejlesztett, Personogram névre hallgató eljárás képes szövegekből emberi tulajdonságokra, személyiségvonásokra, érzelmi állapotokra következtetni. Ez egyfajta digitális profilozó módszer, amely személyiségvonásokról és azok összefüggéseiről képes információt szolgáltatni szöveges adatokból.

Egyedi vonástérképet kapunk további statisztikai feldolgozások eredményként, amely több, mint húsz személyiség tulajdonságról képes számot adni. Képet kapunk arról is, hogy az adott szövegben milyen érzelmi viszonyulások jelennek meg, milyen idői orientációt tartalmaz, azaz a szöveg inkább a múltra, jelenre, jövőre reflektál-e. A profilozás eredménye mentén lehetőség van arra, hogy az emergens vonásokat kiemeljük, vagy más profil eredményekkel összehasonlítsuk.

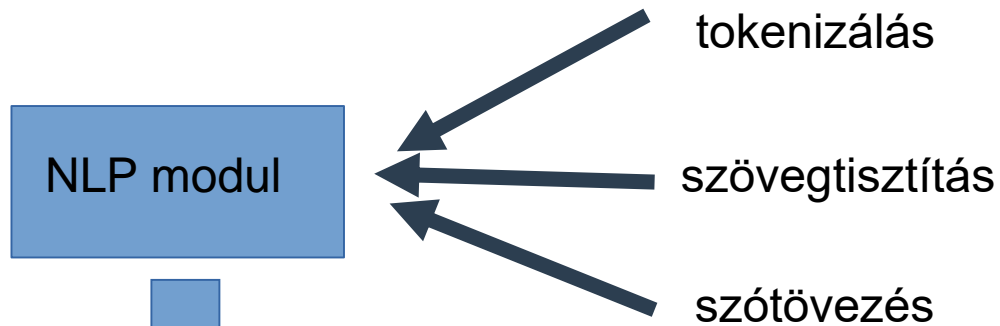
A Personogram rendszer sematikus működése



A Personogram rendszer sematikus működése



Personogram



NLP modul

tokenizálás

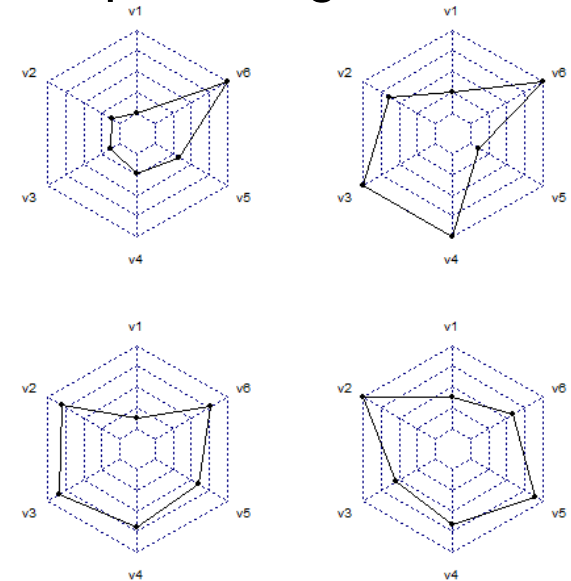
szövegtisztítás

szótövezés

Entity Topic analízis

Illesztésvizsgálat Speciális szótárak
(angol és magyar nyelven)

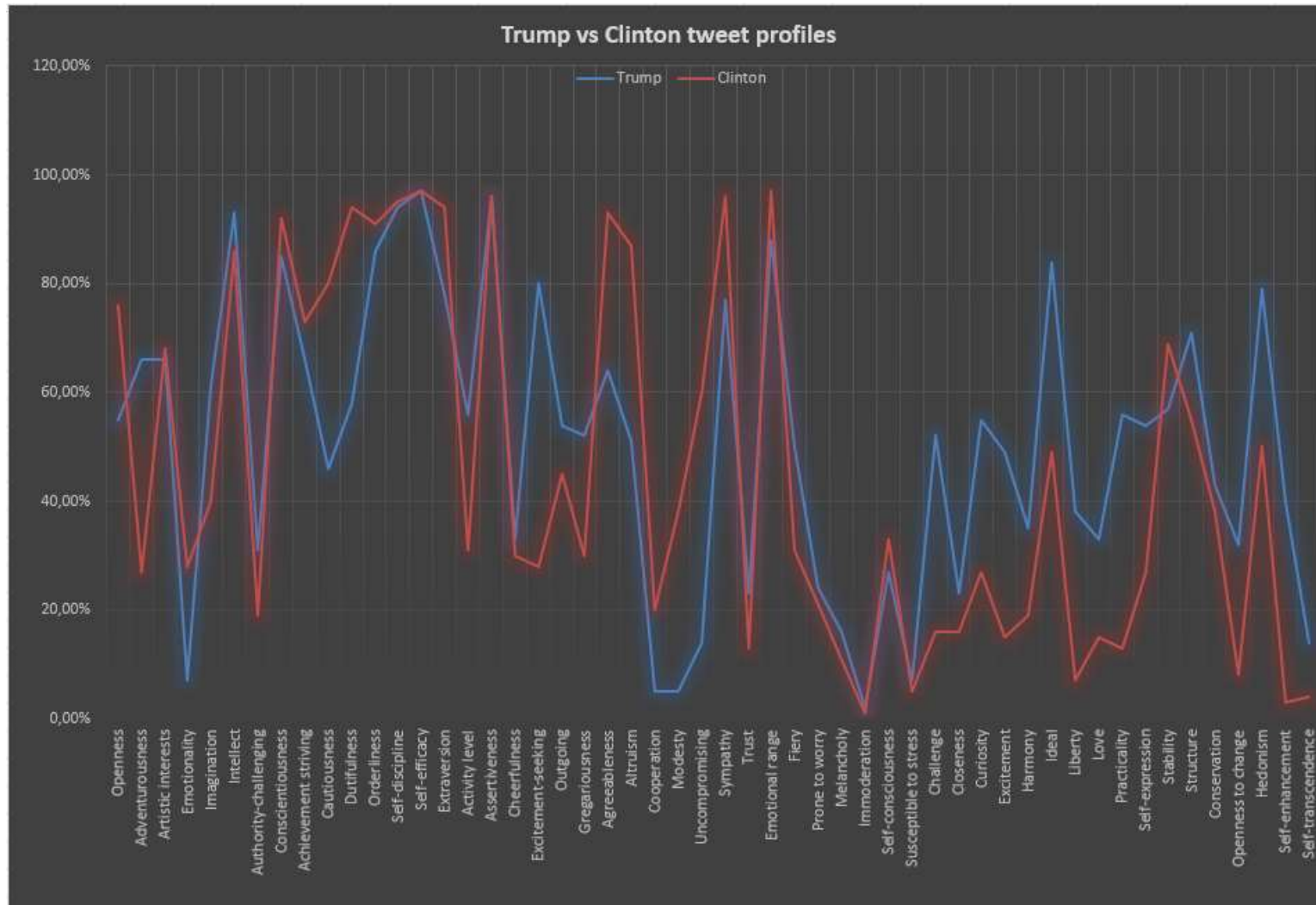
Entitás-alapú
personogram



Példák



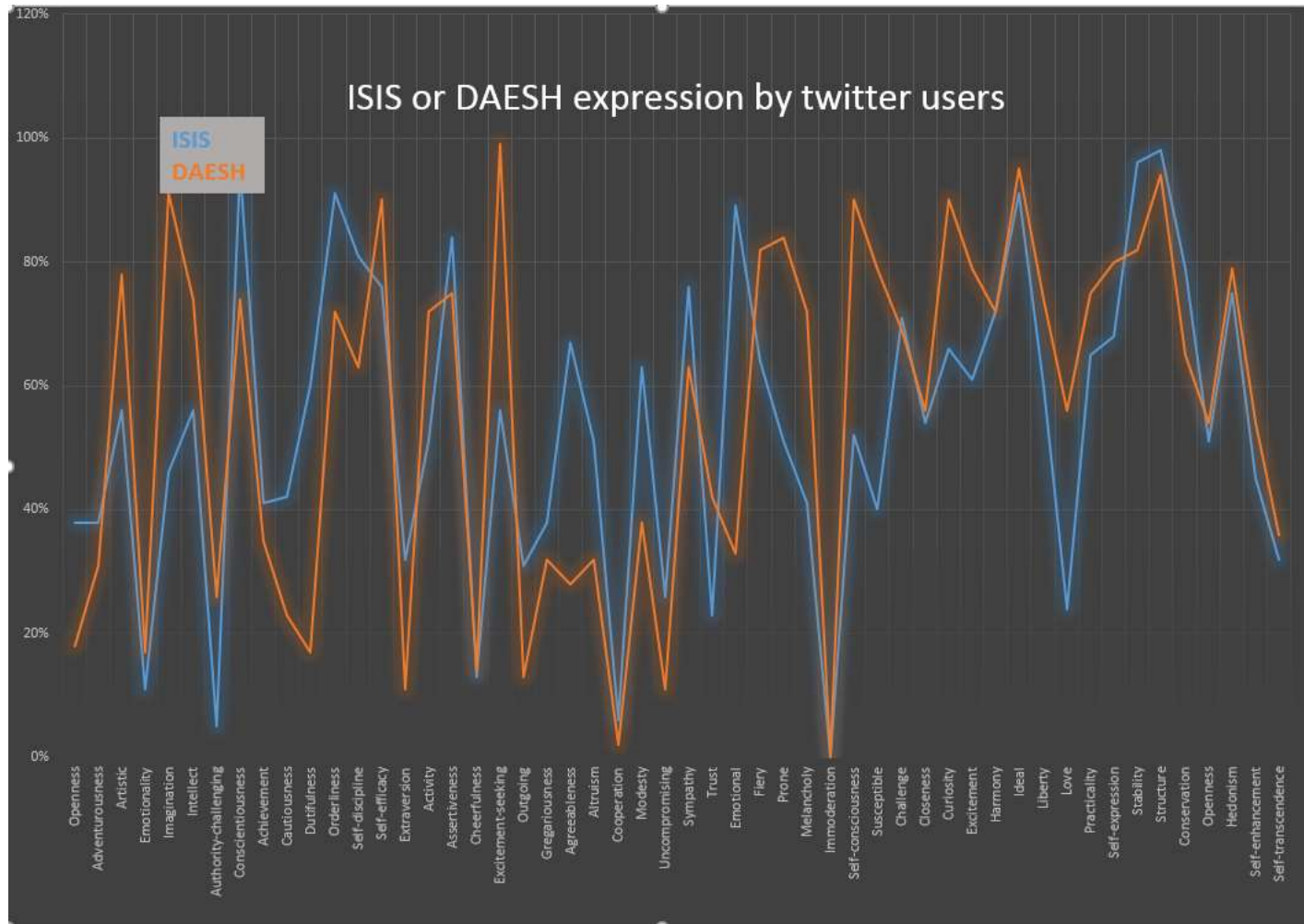
Personogram



Példák



Personogram

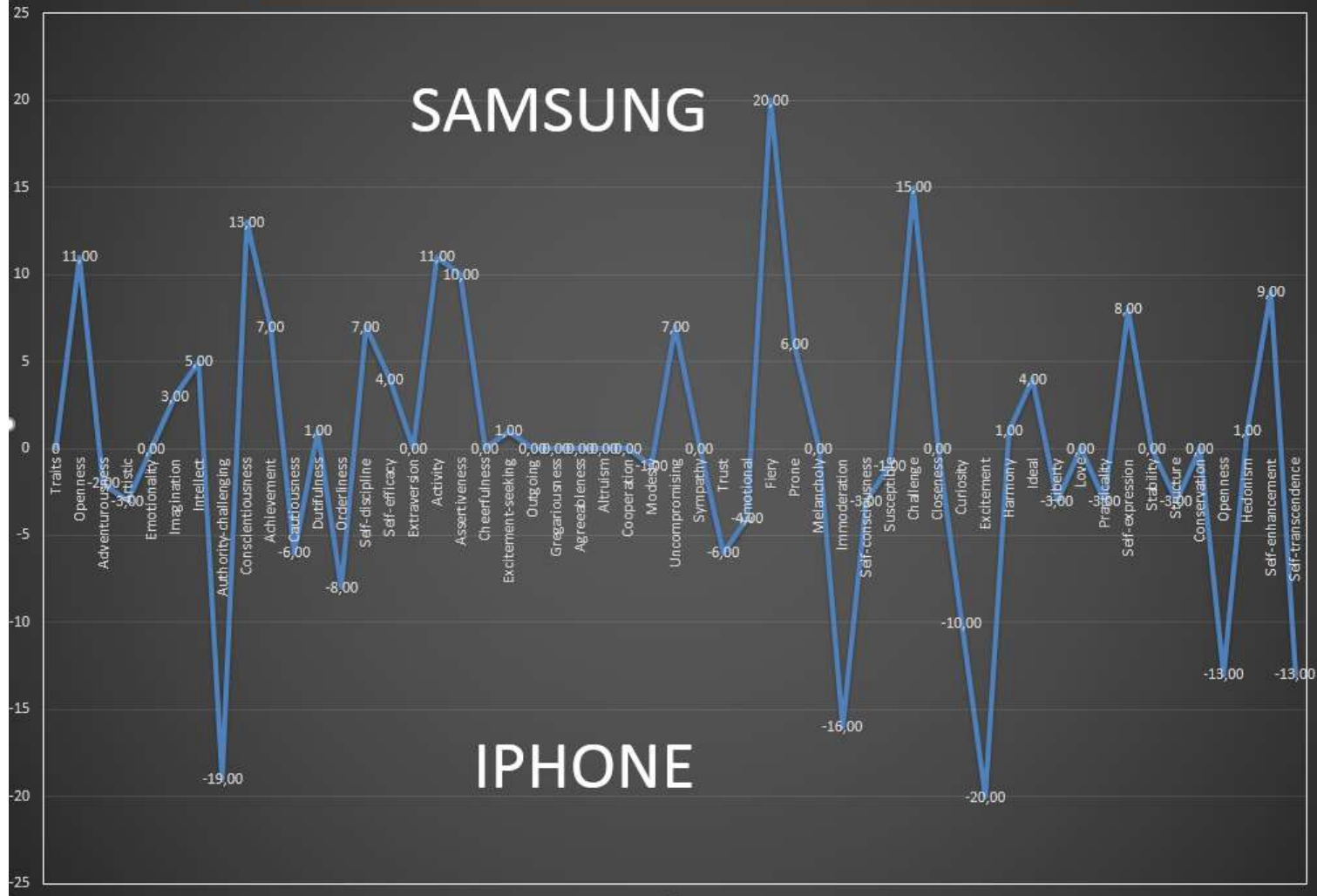


Példák



Personogram

Brand differences between Samsung and Iphone personograms





Personogram

Köszönöm a megtisztelő
figyeletemet!