

NeuroCourt – Anonymisation of Polish Common Court Decisions

1st EU Webinar anonymisation and pseudonymisation
of judicial decisions

Cezary Dołęga, CTO - Neurosoft Sp. z o.o., <cezary.dolega@neurosoft.pl>
29th of May, 2021 - WebEx



Agenda

- Company
- Evolution
- System
- Conclusions

Company



Motto

“ **Nothing is
as practical
as a good theory** ”



Kurt Lewin, 1890-1947

Artificial Intelligence

- **What isn't AI?**
 - Programming
 - Logical reasoning / Rules / Expert systems
 - Databases / Dictionaries
 - Robots / Chatbots / IoT / Drones
- **What is AI?**
 - Extracting knowledge from data / Generalization → **Machine Learning**
 - Knowledge transfer → **Transfer Learning**
 - Artificial Neural Networks → MLP, CNN, RNN, DNN, **GPT, BERT**

Technologies *since 1992*



Speech Processing

Text-to-speech, speech recognition



Digital Image Processing

Optical Character Recognition, Object Detection, Classification & Tracking



Natural Language Processing

Morphology, Syntax, Semantics, NER, NEL, Language modeling



Artificial Neural Networks

Spatioignitron, Deep Learning

Product lines



NeuroLex

Acquisition and processing of legal information

NeuroVoice

Speech and language processing

NeuroCar

Intelligent Transportation System

NeuroSpace

UAV, UGV, UUV – Infrastructure Monitoring

NeuroMedi

Medical Diagnostics

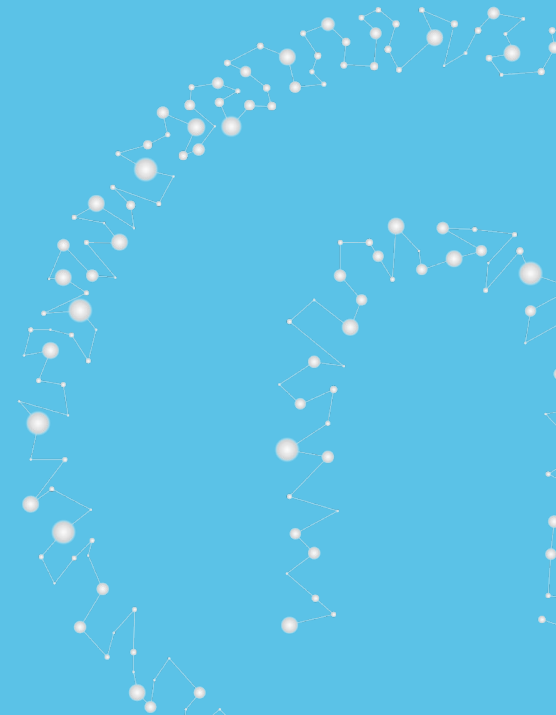
NLP Products

- **NeuroGram**
 - Text tokenisation, stemming, lemmatization, PoS tagger
 - Supports rich inflection of the Polish language
 - Supports dedicated dictionaries (NER)
 - Supports DSL (Scala)
- **NeuroScope**
 - “Language aware” full-text-search
 - In-memory operation
 - Many practical functions for WEB



neurosoft

Evolution



NeuroCourt – Evolution

- Goal ~2010
 - Access to the content of Polish Common Court judgments online
- Challenges
 - Anonymisation
 - Polish Language → inflection, parsing
 - Access to sensitive data → dispersion / scale / security
 - Semantics → structure, taxonomy, identifiers, links ECLI?

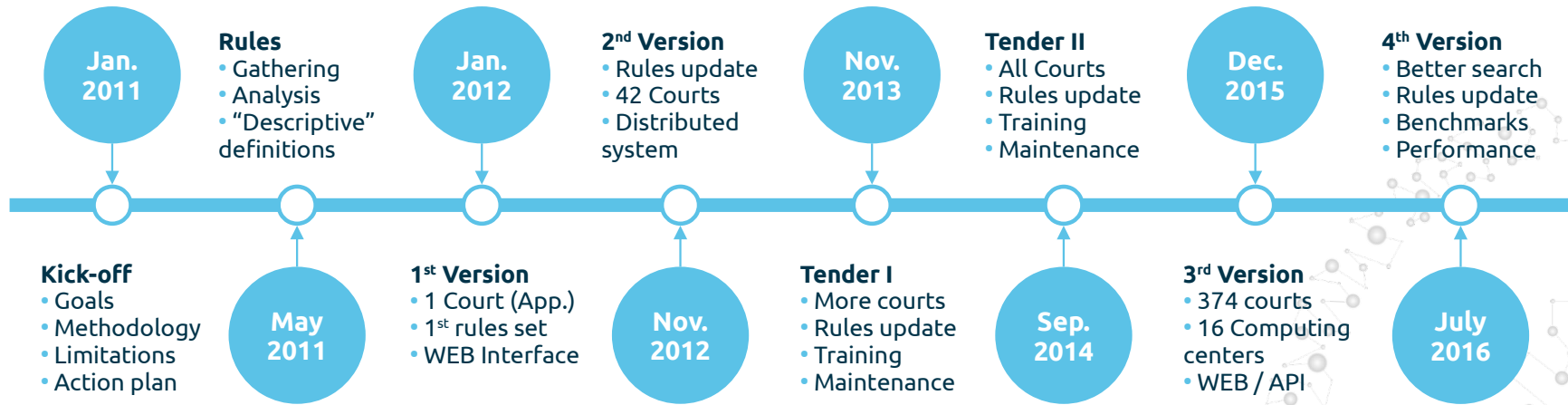
NeuroCourt – Technologies

- OCR / page structure discovery
- Document filtering (DOC, PDF, ...)
- Document structure discovery → XML
- Text tokenization / inflection stemming → **NeuroGram**
- PoS tagging / disambiguation → **NeuroGram**
- Text parsing / phrase structure
- NER / NEL (using Scala/DSL)
- Effective full search → **NeuroScope**

NeuroCourt – Resources

- Legal Acts Database (PL/EU) → titles + metadata + XML
- Patents (PL) → titles + metadata
- Company names (PL) → metadata
- Names of persons (related to companies)
- Geographical places / names

NeuroCourt – Timeline



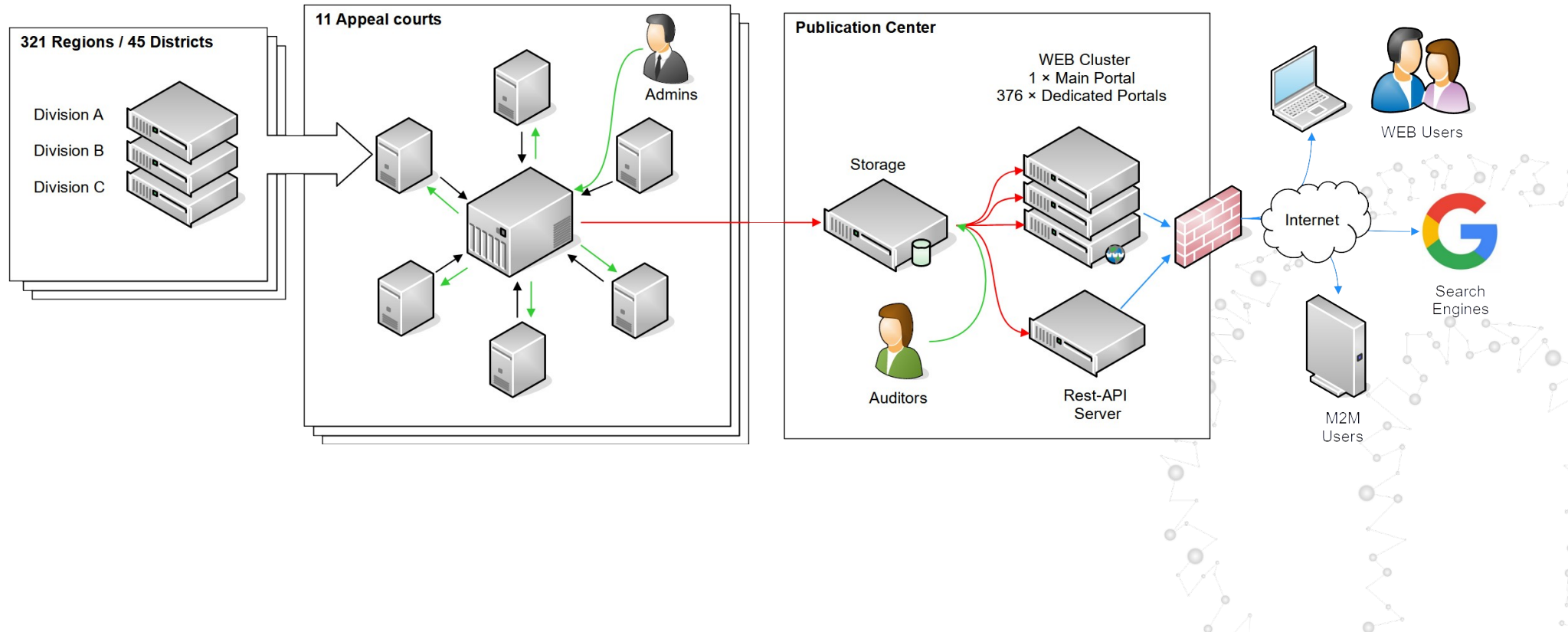


neurosoft

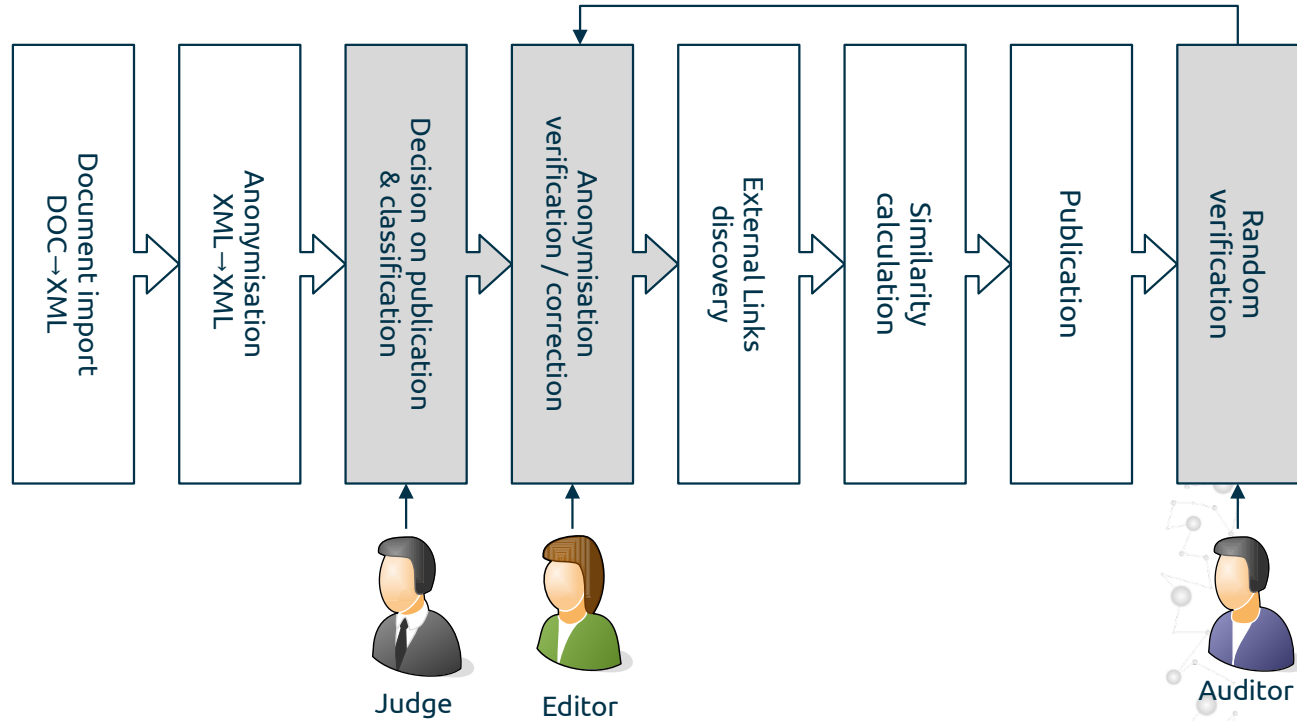
System



Architecture

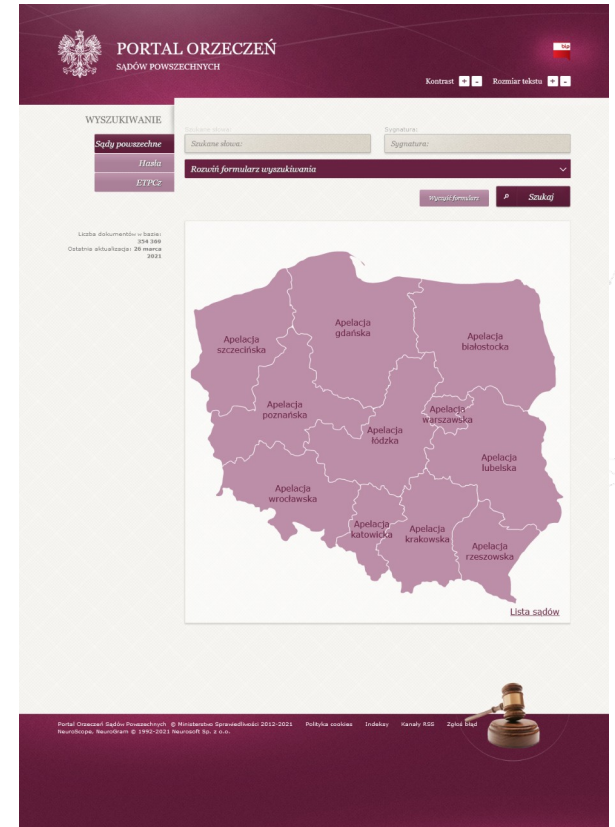


Process



WEB Portal

- <http://orzeczenia.ms.gov.pl/>
- 354k+ documents on-line
- ~2% documents published
- WEB / API access
- Daily update
- On-line support



- Title / signature
- Metadata
- Content
- Legislation
- Similar judgments
- Important terms

VI ACa 564/12 - wyrok z uzasadnieniem Sąd Apelacyjny w Warszawie z 2012-11-23

Metryka Treść Powołane przepisy Orzeczenia podobne

Tytuł: Sąd Apelacyjny w Warszawie z 2012-11-23

Data orzeczenia: 23 listopada 2012

Data publikacji: 25 maja 2013

Data uprawomocnienia: 23 listopada 2012

Sygnatura: VI ACa 564/12

Sąd: Sąd Apelacyjny w Warszawie

Wydział: VI Wydział Cywilny

Przewodniczący: Ewa Stefańska

Sędziowie: Maciej Kowalski

Protokolant: sekr. sąd. Beata Pelikańska

Hasła tematyczne: Darowizna

Podstawa prawna: art. 898 § 1 k.c.

Teza: *Darczyńca może odwołać darowiznę nawet już wykonaną, jeżeli obdarowany dopuścił się względem niego rażącej niewdzięczności. Odwołania darowizny nie uzasadnia jednak każde naruszenie przez obdarowanego norm obowiązujących w społeczeństwie. Jak słusznie stwierdził Sąd Okręgowy, rażąco niewdzięczne zachowanie obdarowanego (działanie lub zaniechanie skierowane bezpośrednio lub nawet pośrednio przeciwko darczyńcy, które, oceniając rzecz rozsądnie, musi być uznane za wysoce niewłaściwe i krzywdzące darczyńcę) musi dotyczyć bezpośrednio darczyńcy. Ewentualnie, cech rażącej niewdzięczności mogą mieć także zachowania wysoce nieprzyjazne, kierowane do osoby bliskiej darczyńcy z zamiarem oddziaływania bezpośrednio na darczyńcę. A zatem niewdzięczność wobec osoby bliskiej darczyńcy, bez związku z jego osobą, nie uprawnia do czynności, o których mowa w ww. przepisie.*

VI ACa 564/12 - wyrok z uzasadnieniem Sąd Apelacyjny w Warszawie z 2012-11-23

Metryka Treść Powołane przepisy Orzeczenia podobne

Sygn. akt VI ACa564/12

WYROK

W IMIENIU RZECZYPOSPOLITEJ POLSKIEJ

Dnia 23 listopada 2012r.

Sąd Apelacyjny w Warszawie VI Wydział Cywilny w składzie:
Przewodniczącej Sędzia SA – Ewa Stefańska
Sędzia SA – Maciej Kowalski
Sędzia SA – Ksenia Sobolewska – Filek (spr.)

Protokolant: – sekr. sąd. Beata Pelikańska

po rozpoznaniu w dniu 23 listopada 2012r. w Warszawie
na rozprawie sprawy z powództwa T. S.
przeciwko M. B.
o zobowiązanie
na skutek apelacji pozwanego
od wyroku Sądu Okręgowego w Warszawie
z dnia 22 grudnia 2011r.
sygn. akt XXV C 1713/09

I zmienia zaskarżony wyrok w całości ten sposób, że oddala powództwo T. S. obowiązkiem zwrotu kosztów procesu na rzecz M. B.;

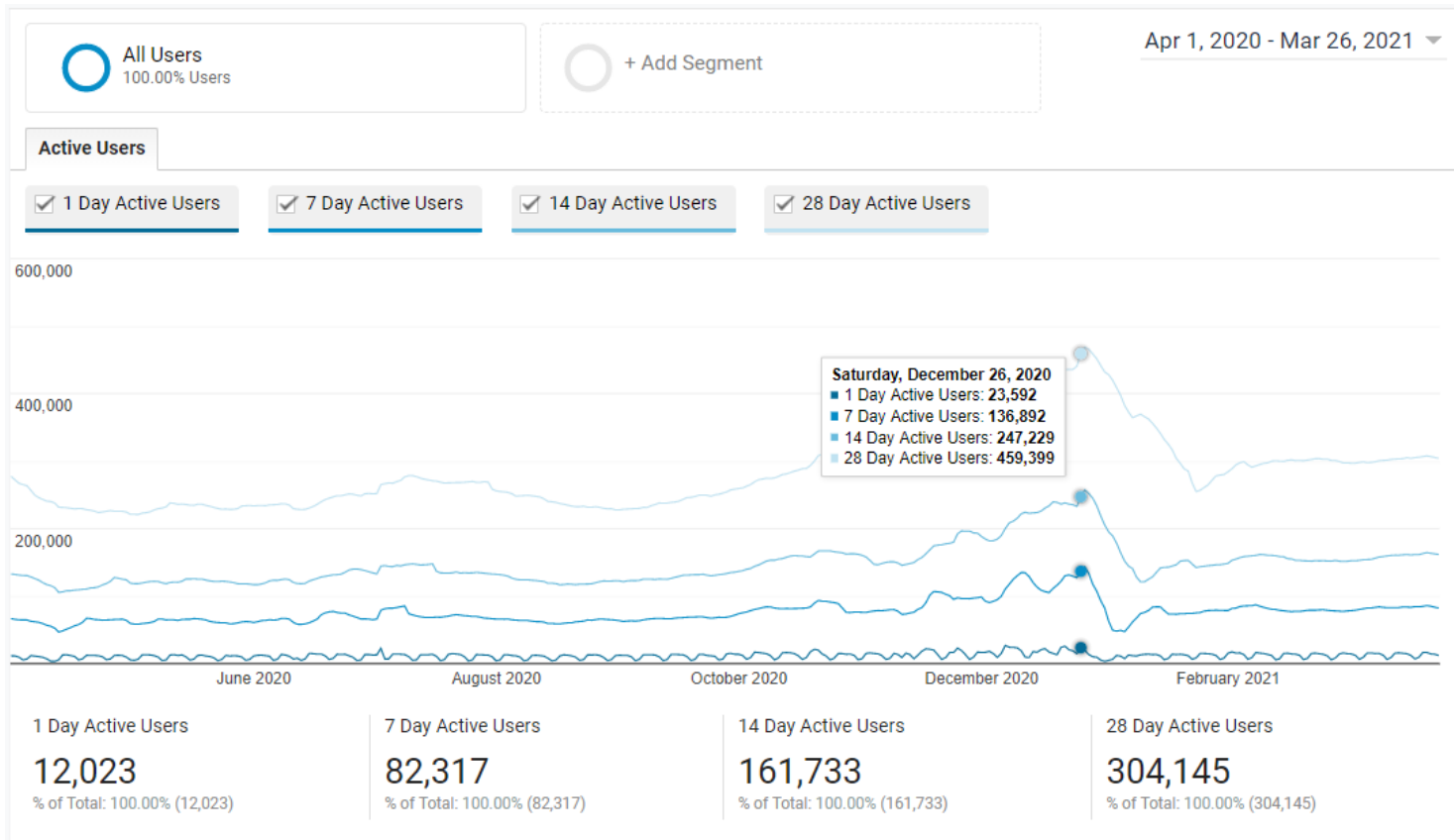
II nie obciąża T. S. obowiązkiem zwrotu kosztów procesu na rzecz M. B. w instancji odwoławczej.

Sygn. akt VI ACa564/12

UZASADNIENIE

Wyrokiem z dnia 22 grudnia 2011r. Sąd Okręgowy w Warszawie, uwzględniając powództwo T. S. przeciwko M. B., zobowiązał pozwanego do złożenia oświadczenia woli, moq którego pozwany

WEB Portal – Usage



Anonymization reliability

- **Q: How often does automatic anonymization fail?**
- **A: Depends on the assessment methodology...**



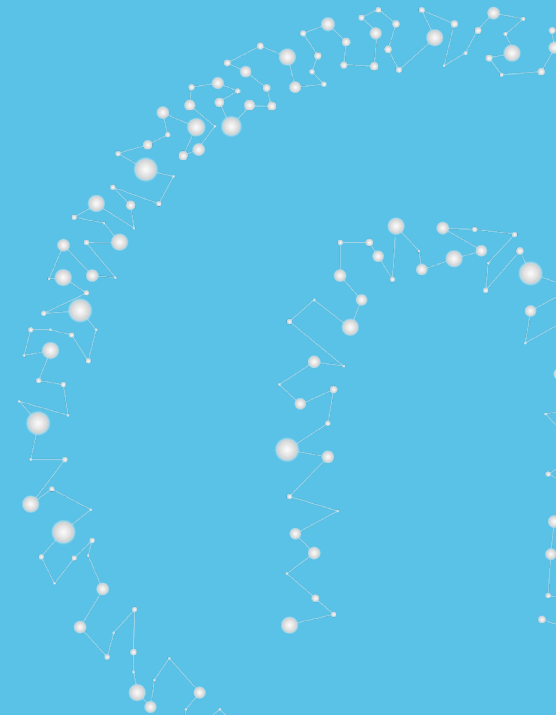
- Segment / sentence level

- high accuracy > 97%
- a lot of unit tests
- benchmarks

- Document level

- Multiple elements per document >20
- $0.97^{20} = 0.54 \dots$
- better accuracy is required

Conclusions



✓ Achievements

- Citizens have access to case law (since 5 years)
- Access to the content of judgments immediately / on-line
- The system works in all common courts → “full scale”
- Automated anonymisation, consistent across all courts
- Extensive experience with anonymisation problems

✘ Failures

- Only 2% of the judgments published
- Reliability of automatic anonymisation needs improvement
- Manual work should be reduced
- Little use of contemporary AI technologies

Future objectives

- **Knowledge sharing within the European Union**
 - International cooperation is essential
 - Poland joins ECLI?
 - Common rules for anonymisation / common tools / benchmarks ?
- **Better utilization of Artificial Intelligence**
 - Using **BERT** model to improve the quality of anonymisation (NER, de-anonymisation assertions)
 - Creation of a language models based on multilingual EU data
 - Applying AI to search and reasoning – "**questions answering**"



neurosoft

Artificial Intelligence

– let's put “good theory” into practice



Thank you !

