



Трети ден, 7 октомври 2020

ИНИЦИАТИВИ НА ГОДИШНАТА КОНФЕРЕНЦИЯ НА КЛАРИН 2020

ПАНЕЛ ЗА ИЗКУСТВЕН ИНТЕЛЕКТ, ЕЗИКОВИ ДАННИ И ИЗСЛЕДОВАТЕЛСКИ ИНФРАСТРУКТУРИ (PANEL ON ARTIFICIAL INTELLIGENCE, LANGUAGE DATA AND RESEARCH INFRASTRUCTURES)

Уебинарът представя четирима панелисти - **Vikusi Marivate**, **Jan Hajic**, **Malvina Nissim**, **Sien Moens**. Публиката може да задава въпроси към тях.

Първият въпрос беше с коя научна област най-много се идентифицират: изкуствен интелект, дигитална хуманитаристика, компютърна лингвистика, обработка на естествен език, компютърни науки.

Malvina Nissim каза, че работи предимно с изкуствен интелект и език, но понеже те са дълбоко свързани с други области, може да се каже, че се идентифицира с всички изброени.

Sien Moens обясни, че се занимава предимно с обработка на естествен език, но работи и по техники за машинно обучение, които пък са свързани с изкуствения интелект и компютърните науки. Интересува се също от дигитална хуманитаристика и компютърна лингвистика.

Jan Hajic отговори, че е работил във всички изброени области, но по различно време. Работата му в DARIAH и CLARIN е свързана и с дигитална хуманитаристика.

Vukosi Marivate каза, че е започнал с изкуствен интелект, но по-късно е работил в областта на компютърните науки, а през последните пет години е съсредоточен върху обработката на естествен език. Анкетата на участниците показва, че най-голям брой от тях се идентифицират с областта на дигиталната хуманитаристика (30%).

Вторият въпрос към панелистите се отнасяше до връзката им с CLARIN.

Malvina Nissim каза, че не е официално свързана с CLARIN, но е потребител на ресурсите, а заедно с колегите си в университета предоставят ресурси.

Jan Hajic е официален член - координатор за Чехия.

Vukosi Marivate е част от институцията, която е свързана с CLARIN, определи се като наблюдател и потребител.

Sien Moens каза, че участва редовно в събитията на CLARIN и, че наскоро тя и екипът ѝ са изпратили предложение за участие и се надява да бъдат одобрени.

Имаше въпрос към **Jan Hajic**: защо хората трябва да имат интерес към CLARIN?

Той отговори, че участието в CLARIN може да донесе само позитиви - да получаваш подкрепа от общността, да имаш достъп до ресурси, инструменти, които са лесни за намиране и много добре обяснени. Каза, че участието в CLARIN дава възможност да направиш много за екипа и държавата си, включително и във финансов аспект.

Следващият въпрос към **Malvina Nissim** беше: от какво има нужда езикът в областта на изкуствения интелект? Тя каза, че има много опит със събирането на данни и в тази сфера на работа достъпността на данните е сериозен проблем. Напоследък по-голям проблем е не намирането на данни, а начинът на използване. Изследователите в тази област се сблъскват с въпроса: позволено ли е да събирам, съхранявам или споделям данни?

Каза, че правните аспекти са много трудни, че липсва яснота по темата. Би се радвала, ако CLARIN стане център за поставяне на добри практики в това отношение, ако помага на изследователите да разберат какво е позволено и какво не е. Каза, че в наши дни изследователите искат да споделят данни по целия свят, а законовите рамки за това се решават на локално ниво от отделни хора, липсва глобално решение.

Следващият въпрос беше към **Vukosi Marivate**: работиш с големи данни, имаш опит с езици с малко ресурси, каква е тяхната роля в областта на изкуствения интелект, предвид доминацията на английския език?

Vukosi Marivate отговори, че интернетът става едноезичен, а хората искат достъп до информация на своя език, нуждаят се от преводи, затова трябва да се работи в тази посока.

Sien Moens говори за предизвикателствата на изкуствения интелект в социален аспект. Каза да си представим как бихме говорили с робот или роботизирана кола без шофьор. Човек би говорил по много естествен начин, както говори с останалите хора, а всъщност трябва да съобрази говора, възприятията си с изкуствения интелект. Изследователите в тази област трябва да комбинират знания за аудио-визуалните възприятия и езика, за да интегрират човешкото разбиране за естествения език; трябва да намерят начин за интегриране на общата култура, която осигурява общ фон от знания между хората. Даде пример с това, че, ако искаме да накараме колата без шофьор да спре до някое дърво, за да вземем приятел, трябва да съобразим какви са възприятията на колата и как да представим посланието си, за да го разбере. Заклучи, че са необходими много ресурси за развитие в тази посока.

Darja Fišer попита каква е ролята на CLARIN в областта на изкуствения интелект?

Jan Hajic каза, че подкрепя твърдението на **Sien Moens** за нуждата от повече мултимодални данни. Допълни, че CLARIN е насочена основно към езика, но е време да задълбочи и да направи по-силна връзката си с изкуствения интелект. Каза, че има мрежи от хора, които работят в тази област, и те също събират данни за езика. Затова трябва да се изгради интердисциплинарен подход и да се работи екипно в тези области.

Malvina Nissim изтъкна проблема за техническата инфраструктура - не всичко се свежда до намиране на данни, необходима е добра техническа база. Липсата на такава затруднява създаването на ресурси и инструменти. Надява се CLARIN да може да помага в това отношение.

Последният въпрос беше към **Jan Hajic** - каква е основната полза от CLARIN?

Той отговори, че основната ѝ роля трябва да е да поддържа езика, но и да работи с други области; да се насочи и към други езици, които може да не се говорят от много хора, но все пак да имат ресурси или обратното - да са големи езици, но с малко ресурси; да се опита да работи извън Европа; да работи с експерти от други сфери - изкуствен интелект, психология и други.

IMPROBOTICS - ИМПРОВИЗИРАНО ТЕАТРАЛНО ПРЕДСТАВЛЕНИЕ (IMPROBOTICS - IMPROVISED THEATRE SHOW)

В този модул беше показана театрална импровизация с помощта на инструмент за автоматично разпознаване на реч - актьори, говорещи на различни езици (полски и нидерландски) импровизираха заедно като разчитаха на автоматичен превод. По-късно импровизацията включи и чатбот, основан на изкуствен интелект - един от актьорите произнасяше реплики, генерирани от чатбота. В последната част актьорите импровизираха със стихове, чатботът също генерираше такива. Авторите предоставиха запис на представлението: <https://www.youtube.com/watch?v=XJxyv3BxPil>

СТУДЕНТСКА СЕСИЯ НА CLARIN (CLARIN STUDENTS SESSION)

Maciej Maryl обяви тази сесия за една от най-важните в конференцията, защото студентите и докторантите са бъдещето на CLARIN.

Laura Picchio представи корпус от данни за филмови фестивали - двуезична реч (италиански и английски език), аудио и видео материали.

Anna Medrzecka представи проект за иронията в полската литература и особено в творчеството на Juliusz Słowacki. Дигитализирала е всички текстове, за да състави корпус. Използвала е инструмента за честотни списъци на CLARIN - LEM, разпознаване на именувани същности (NER) и WordNet.

Harriet Cliften показва лексикален ресурс за модели на знание в културното наследство. В момента работи по първия етап - създаването на класификация на културно наследство.

Zaida Batolome-Diaz представи френско-испански онлайн дигитален речник, специализиран в археологията. Данните за корпуса са събрани от интернет с инструментите на SketchEngine.

Iveta Kopankina показва ресурс от резюмета на дисертации, преведени от английски на латвийски. **Alina Tsikulina** представи многоезичен паралелен корпус.

Mateusz Gniewkowski показва проект за оценяване на качеството на алгоритми.

CLARIN В КЛАЧАТА СТАЯ (CLARIN IN THE CLASSROOM)

Francesca Frontini каза, че ползите от тази инициатива за CLARIN са изграждането на силна общност от учители и лектори, а ползите за останалите са ресурсите, инструментите, възможностите за колаборация и подкрепа.

Mietta Lennes представи проект за построяване и поддържане на онлайн курсове по методи за дигитални изследвания.

Laura Herzberg представи курс за използване на инструменти и платформи за корпуси, предназначен за университети.

Jurgita Vaicenoniene говори за педагогическите приложения на корпуса ORVELIT с оригинални и преведени на латвийски текстове.

Kristina Pahor представи ръководство за използване на парламентарни корпуси, подходящо за различни типове специалисти.

Federico Boschetti говори за различни семинари на CLARIN за студенти, проведени в четири университета и две гимназии в Италия.

Simonetta Montemagni представи курс за студенти по обработка на естествен език. **Inguna Skadina** говори за инициативата CLARIN in classroom в Латвия.

Mikko Tolonen говори за използването на данни и инструменти на CLARIN в областта на дигиталната хуманитаристика.

Lonneke van der Plas представи проекта UPSKILLS и финансираното от Erasmus+ сътрудничество на пет университета и CLARIN ERIC.

Последните модули на конференцията бяха [CLARIN Bazaar](#) и индивидуални срещи с докторанти и преподаватели, разделени в отделни виртуални стаи, където желаещите могат да получат повече информация за работата им и да задават въпроси.

Конференцията беше закрыта от **Monica Monachini** и **Liané Van Den Bergh**, които изказаха благодарности към всички, които са допринесли за осъществяването на конференцията в този нов виртуален формат, към всички участници и към публиката.

ПОЛЕЗНИ ЛИНКОВЕ:

- <https://www.clarin.eu/governance/legal-issues-committee>
- <https://www.clarin.eu/content/legal-information-platform>
- <https://nsd.no/clarin-k-centre>
- <https://arxiv.org/abs/2010.02353>
- <https://www.masakhane.io>
- www.improbabilities.org
- <https://github.com/kingsdigitallab/sdlc-for-rse/wiki/Data-Management-Plan-guidance-and-AHRC-template>

Подготвил: Зара Кънчева, ИИКТ-БАН