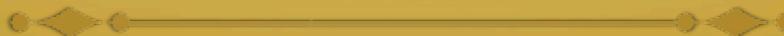
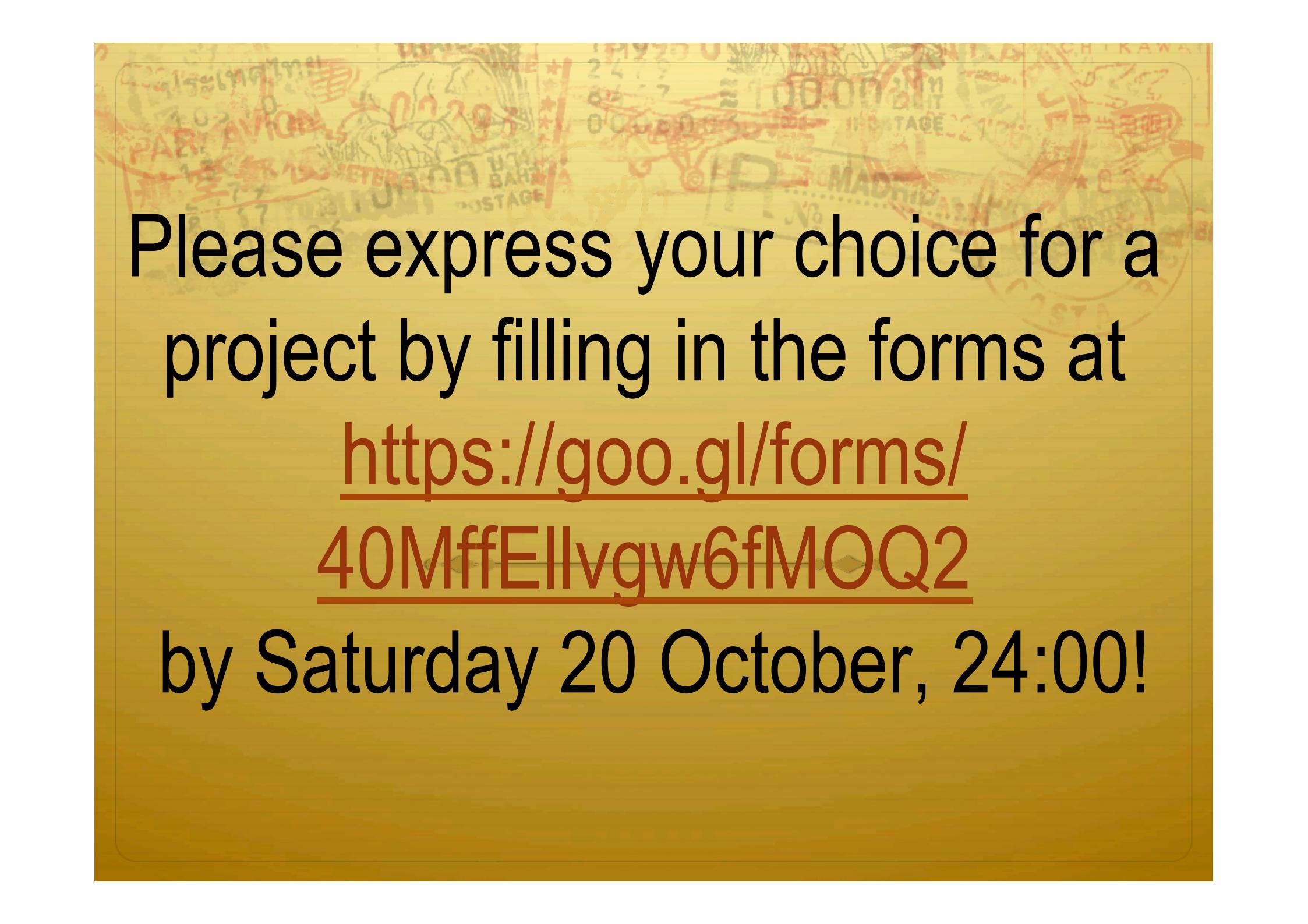


Course 2

IA projects list



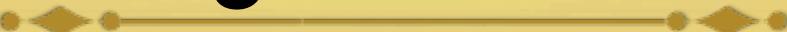


Please express your choice for a project by filling in the forms at

[https://goo.gl/forms/
40MffElvgw6fMOQ2](https://goo.gl/forms/40MffElvgw6fMOQ2)

by Saturday 20 October, 24:00!

How will the activity be organised?



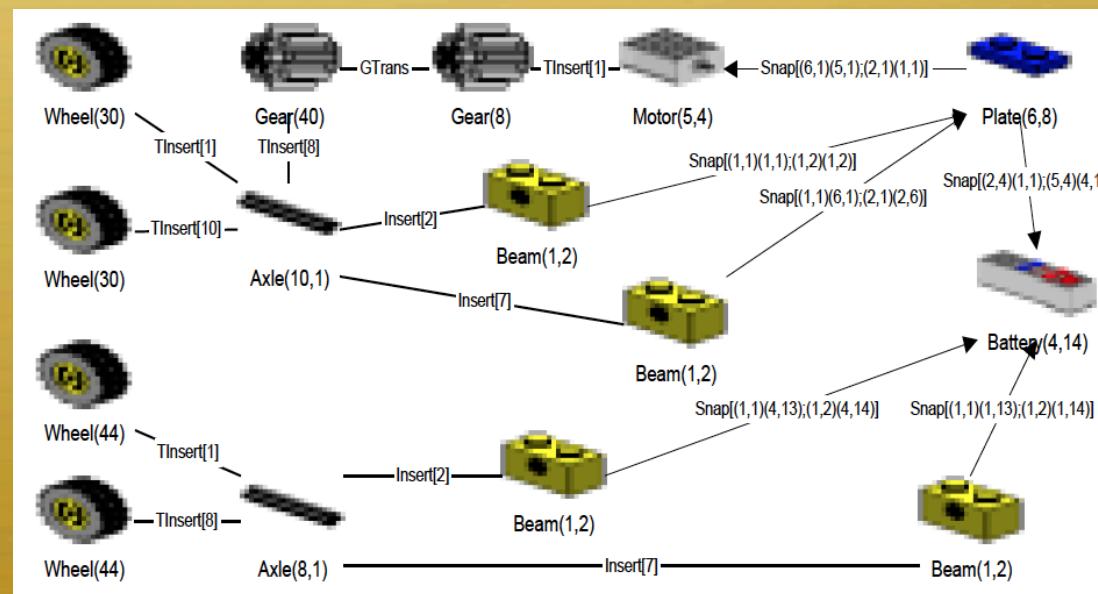
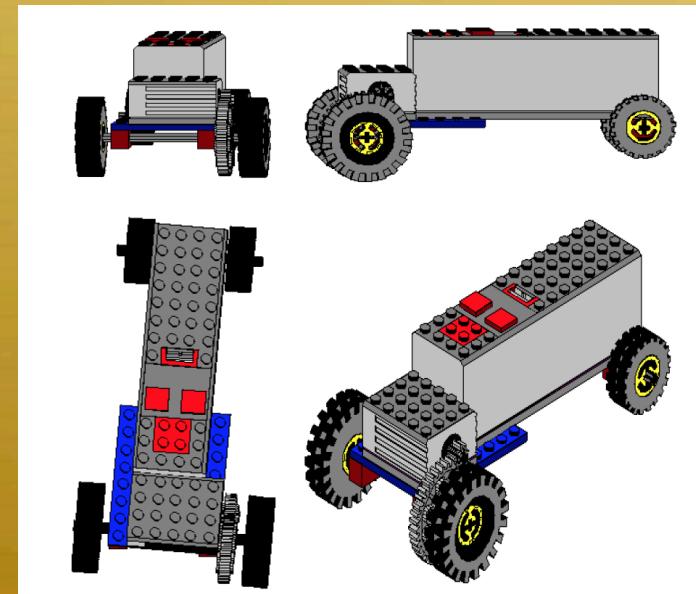
- ◆ Each student votes for 3 projects, in descending order of liking
- ◆ Votes are cumulated at the level of student groups (teams)
- ◆ In each team, the project with the majority of likes will be selected
- ◆ Restrictions:
 - ◆ for competitiveness, at least two groups take the same project
 - ◆ possibly, a superior limit also set (perhaps 3)
 - ◆ assistants should, more or less, emphasize with the themes
- ◆ General conceptual issues discussed in classes (courses)
- ◆ Design and realisation details fixed during labs
- ◆ The team proposing the best solution among those assigned to the same project will receive bonuses

P1. Lego assemblings

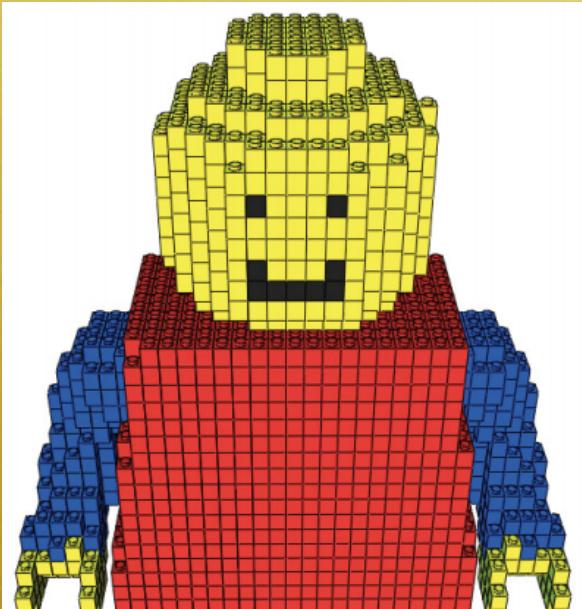


- ◆ Domain: optimisation of topological structures, automatic generation of brick sculptures
- ◆ Method: most likely, genetic algorithms, but others are accepted
- ◆ Suggested bibliography:
 - ◆ Peysakhov, Galinskaya and Regli: Representation and Evolution of Lego-based Assemblies
 - ◆ Funes and Pollack: Automatic Generation of Constructable Brick Sculptures; Computer Evolution of Buildable Objects
- ◆ Representations: assembling graphs, Lego graph grammar
- ◆ Idea
 - ◆ encode the assembly graph as a chromosome
 - ◆ GA encoding scheme:
 - ◆ initialisation, mutation, cross-over
 - ◆ use an evaluation function to estimate fitness of the obtained structure

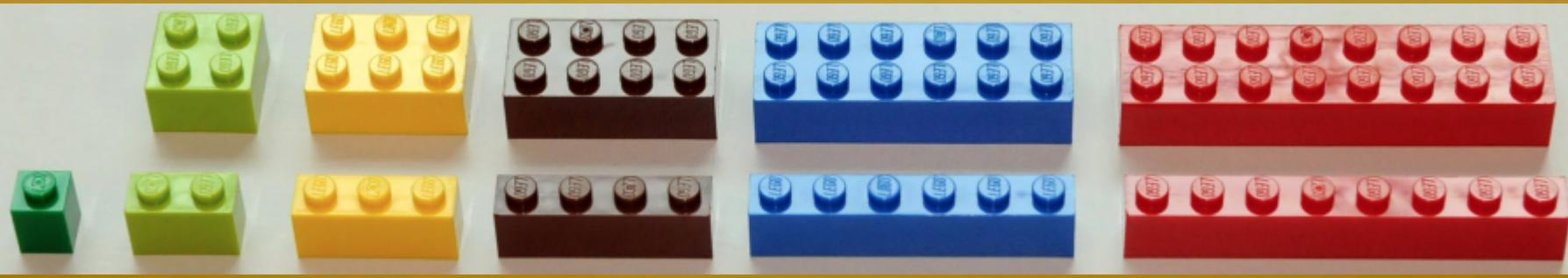
Assembly graph for a Lego car



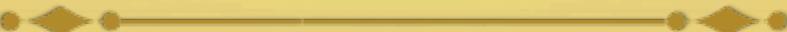
Creating Lego sculptures with bricks



1. fill in the 3D space with 1 x 1 bricks
2. merge 1 x 1 bricks to larger structures
3. solve instability issues
4. decompose into sequential steps



P2. Plant breeding



- ◆ Domain: computational biology
- ◆ Example of AI based applications in agriculture: autonomous tractors, sensor-driven fertilizer applications or robotic weed control; in general, learning from rapidly expanding mega-datasets (Big Data) generated in crop breeding programs
- ◆ Goals: phylogenetic analysis (analyze phylogenetic trees of different organisms), sequence alignment (find motifs or common patterns in a set of given DNA sequences)
- ◆ Methods: Local Rank Distance \Rightarrow given a set of reads that need to be aligned against a reference genome, the aligner determines the position of each read in the reference genome that gives the minimum Local Rank Distance

Plant breeding: what and how?



Goal: predict performance of plant breeding lines and make intelligent selection decisions to estimate their “behaviour” under potential future crop production scenarios

Methods: multiple kernel learning framework, deep AI networks similar to word embedding in natural language processing (Mikolov et al. 2013; Mejia-Guerra and Buckler, 2017)

Local Rank Distance (LRD) aligners



Task: align reads sampled from several mammals on the human mitochondrial DNA sequence genome

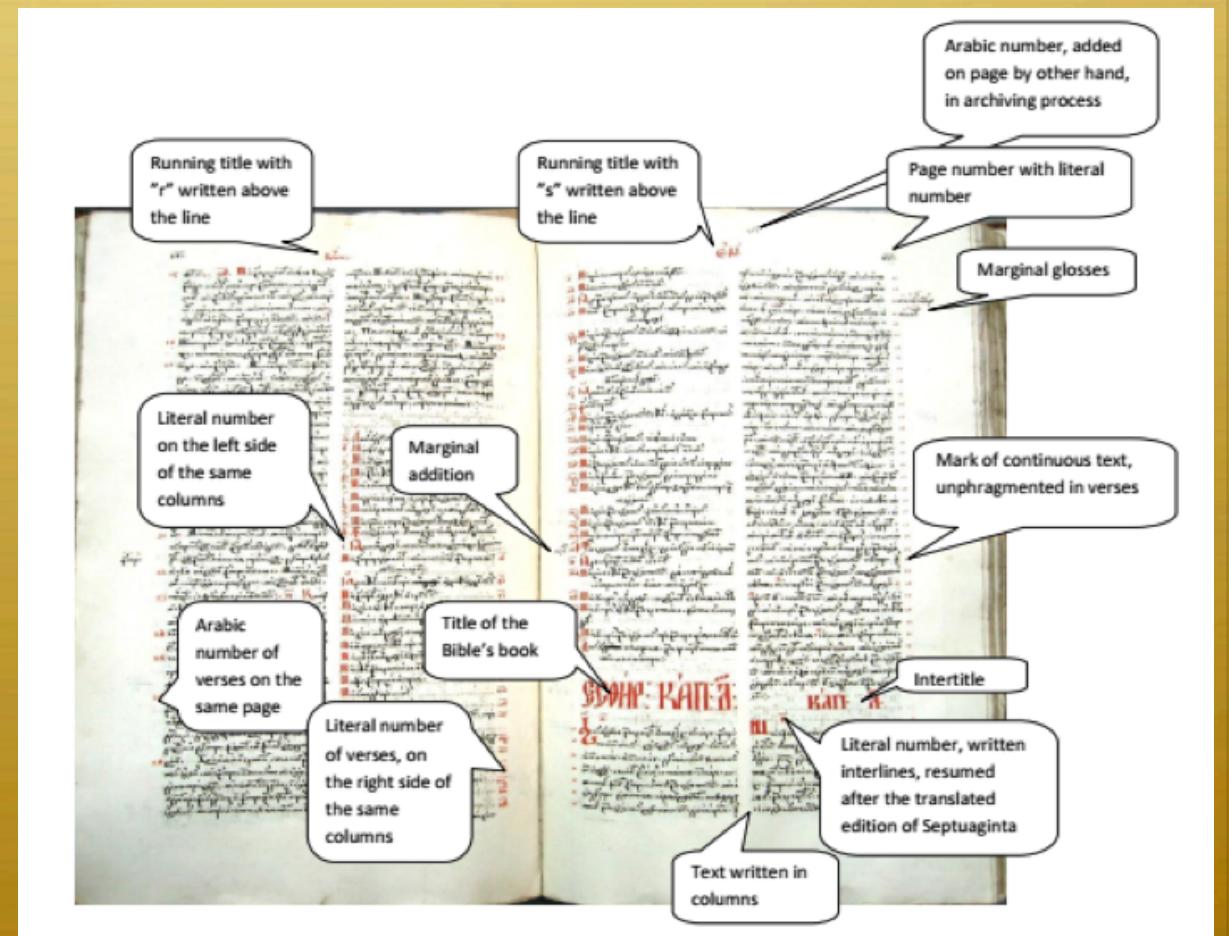
- ◆ maximize the number of aligned reads sampled from the human genome (true positives) and minimize the number of aligned reads sampled from the other mammals (false positives).

Biblio:

- ◆ Iulian Gabur: project application for Marie Curie
- ◆ Radu Ionescu: habilitation thesis ⇒ sequence alignment with Local Rank Distance, pag. 172...
- ◆ Based on: Dinu, Liviu P. and Manea, Florin. An efficient approach for the rank aggregation problem. *Theoretical Computer Science*, 359(1{3}): 455{461, 2006.

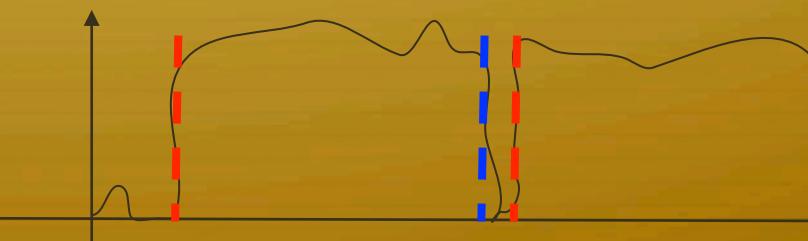
P3. Segmentation of writings: prints and manuscripts

Use learning
algorithms to
decipher positions
of columns, lines,
interlinear writings,
cuff notes, titles,
etc.



How to detect columns?

Project the page image on the horizontal axis and do a histogram of pixels

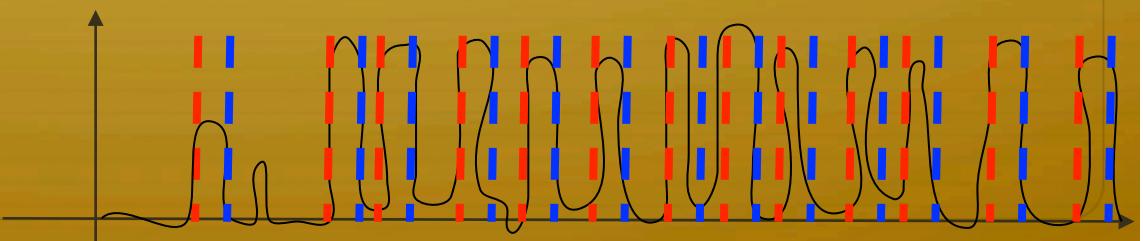


How to detect lines?

For each detected column, project the page image on the vertical axis and do a histogram of pixels

(a)

Линии, проявляющиеся на изображении



How to discover marginal notes?

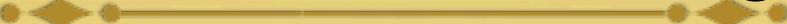
Do the X-followed-by-Y projections for the contiguous zones which are close to zero but not identical to zero \Rightarrow then learn to distinguish marginal notes, headers, footnotes, page numbers from spots, artefacts, etc.

лара, котрж кабіткүйкінгіл ағып Дінегеу сағ
факт . Кәдірланаң чөлшір честағ лаңгы
Дінегеу. Але Зәңбек, сұғағчым шы . 18даң-
екша астың ақаста плахмабыре, шіл Анилеснің

пóате, а́льб. лжепр. шт. лóк не асемнáрб; (п) Шт. а́вфакт Длнезéв (а́дека, оүнвл) пр. шм; Дарнáвфакт бý (а́дека, длнезéй молуý.) Сáв а́парát сфи- та

- (а) Йдикъ, пре Фюл шъ Къвжитъл лъпъ Дънчезъв .
 (б) Йдикъ, прен лъкрабъкъ порънчилшр лъпъ Дънчезъв деславаштъ фъкъндъсъ .
 (г) Йнчъкъ сънтулъ хтодрче къвжитъл кътре йріанъ , кърѣ сънти факъцъ прѣасици жълдовъ лъпъ дъпъ чеъ дин тѣхъ хпротивъторъ лъпъ Хс жълдовъ .
 (д) Бъре : а . г . (е) Кол : а : бї . (з) Ішад : г . Г . (з) Ішад : дї . о .
 (и) Ка йріанъ кърфія ка Фюл лъпъ Дънчезъв не асъменѣ юсте дъпъе фінца къ Татъл .

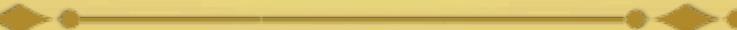
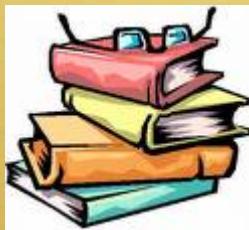
P4. Automatic magazines metadata filling



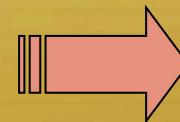
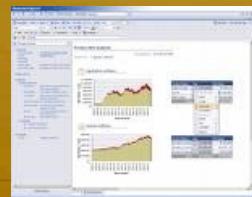
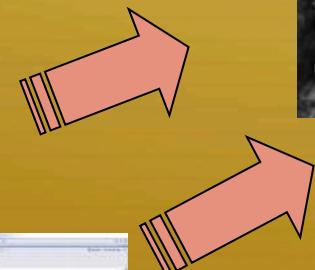
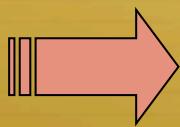
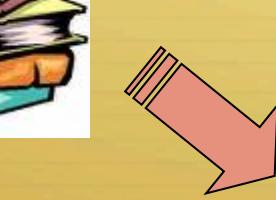
You have a set of triplets: magazine page images + extracted text from the articles + their metadata (title, author, etc.). Write a program able to learn to cut out the texts and their corresponding metadata.

Motivation: building a large corpus or the Romanian language

CoRoLa – achievement of a large corpus of texts



Corpus of
Contemporary Romanian Language



Portal

Curation chain: Portal – Volunteers – Portal

CoRoLa Portal



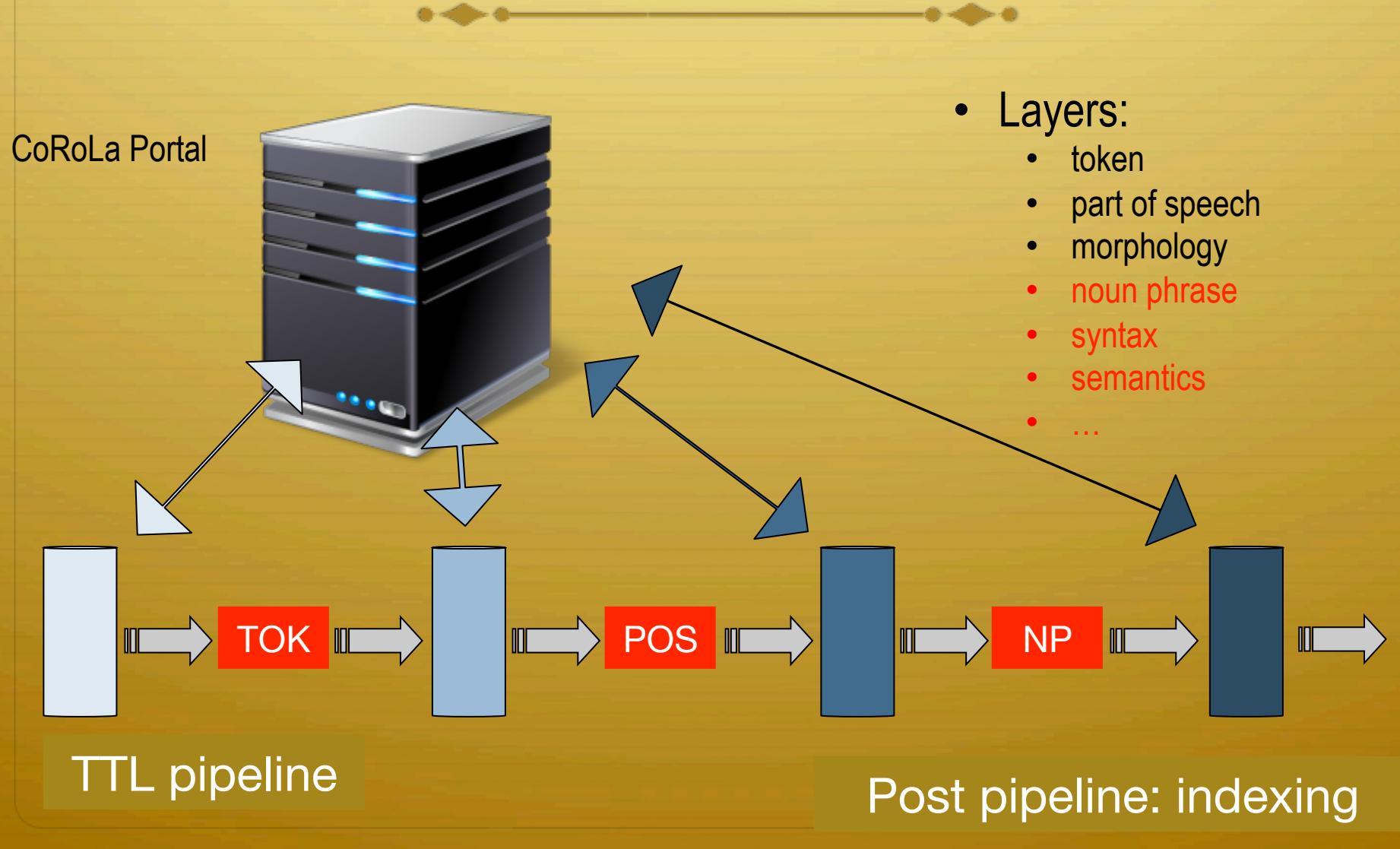
Cleaning

- 1.Character
- 2.Headers
- 3.Footers
- 4.Formulas
- 5.Tables
- 6.Table of contents
- 7.Bibliography
- etc.

Metadata

- 1.Title
- 2.Author
- 3.Publication date
- 4.Source
- 5.Translator
- 6.Media
- 7.Style
- 8.Domain
- 9.ISSN/ISBN

Annotation chain: Portal



Use of the corpus

Acces point:
ICIA MD,
Bucharest

Mirror:
IIT-ARFI,
Iași



programs used in
Computational Linguistics

Concordances

(KWIC – Key Word In Context)

KorAP

in all corpora ▾ with Poliqarp ▾ Glimpse

292 matches [/](#) [/](#)

Corola-blog/BlogPost/375907_a_377236	, și și-ai închis corola -n timpul serii! Tu esti mai albă decât albul iernii, Petalele-ti îmi par a fi de nea, Ai înflorit acum în pragul verii Umplâ
Corola-blog/BlogPost/342573_a_343902	succes Getuța, corola noastră de lumină vie!!! *** Încerc să te descopăr Încerc să te descopăr dintr-o umbră, Chip plămădit în vis din dor
Corola-blog/BlogPost/378264_a_379593	c mălin. Pe sub corola zărilor deschise, ca niște cășile cure de păstori, s-ar aduna, mânate dinspre munte, neîmblânzite turmele de nori. S
Corola-blog/BlogPost/382227_a_383556	ă: „distrugând corola de minuni” a Țării Soarelui Răsirea de bordează de prospețime și originalitate, de o mare încărcătură emoțională și
Corola-blog/BlogPost/347152_a_348481	achii. Unele cu corola rotunda, altele cu frunzele crețe ca de feligă și cu marginile multicolore. Adevărate minunătii ale naturii. Pe marginea
Corola-blog/BlogPost/343751_a_345080	pare își desface corola să prindă între petale toate razele de lumină. Minciu savura cu nesăt senzațiile date de îndemânarea lui Ștefan care-
	ecare popas pe corola florală a unei plante de unde-și culegea dulceața nectarului, ca apoi să pornească în căutarea altor plăceri la fel de d
Corola-blog/BlogPost/350977_a_352306	u gene blânde, corola florilor plăpânde. Le-oi prinde-n piept ca pe-o năframă ce depărtărilor mă cheamă, și-ori energi de cetină albastră voi
Corola-blog/BlogPost/342098_a_343427	eile Zorilor din corola sa solară, lacătele Lacrimilor Înghețate ce au ferecat Porțile Speranțelor ale vistieriei. Praful de stele, adunat de mili
Corola-blog/BlogPost/385158_a_386487	dință este acea corola , floare rară a frumuseții sufletești care crește în adâncurile mistice ale ființei și care este hrănita prin lumenii de lumin
Corola-blog/BlogPost/364559_a_365888	eschid, precum corola unei flori, asupra totalității lumii și presupun transcendența unui absolute. Epoca noastră este epoca autonomizării v
Corola-blog/BlogPost/343521_a_344850	mai închideau corola doar legănarea dansului albastru ne atingea obrajii și cântecul lebedei mai aștepta o poveste -stii de ce se ascund șo
Corola-blog/BlogPost/384216_a_385545	ă ! Și-a strâns corola dalbul ghiocel. - Așteaptă până-i fi mai măricel ! I-am spus eu, vântul, și-am plecat pe-afară: - De-abia-n trei zile fi-
Corola-blog/BlogPost/357890_a_359219	m crești și văd corola cum o îndrepți numai spre mine... și ca un soare bland și cald, prin palme mă strecor în tine, să mă îmbăt la umbra
Corola-blog/BlogPost/360616_a_361945	mului, aşa cum corola își poartă cu gingăsie floarea sa cea mai nobilă. Pe țarina soartei sale s-au afundat împovărații ani - preașindu-i iertă
Corola-blog/BlogPost/363112_a_364441	esc albinele pe corola florilor, culegând cu trompa nectarul din pistil, contribuind astfel și la polenizarea plantelor, ca apoi să zboare mai de

1 2 3 ... 12 ►

~~http://corola.racai.ro~~

Learn from annotated images

Ideea de Universitate

O data cu reformele Bologna, tot mai multe voci pun în chestiune menirea universității. Ce anume ar trebui să pregătească Universitatea? Oameni liberi, buni cetățeni sau forță de muncă? Specialiști cu o calificare înaltă, dar îngrijorători și absenți cu o pregătire mai curând generală? În ce raport ar trebui să se afle cercetarea și predarea? Cu ce scop? În cadrul unei reuniuni organizată de Universitatea cu societate și cu Statul? Tentile care compun acest desfășurare de la Universitatea de vară francofonă, care s-a desfășurat la Institutul Internațional de Studii Politice de la Universitatea „Al. Cuza” din Iași (Stefan Afacan, Bogdan Popescu-Tăriceanu, Mircea Negrușiu), au fost susținute de doi profesori de la Universitatea din Belgrad (Ivan Vuković) și Jan Sokol, profesor de la Universitatea Carol Davila. În 1998, Jan Sokol a fost deputat, semnatar al Charterei 77 și ministru al Educației și Cercetării din Republica Cehă în 1998. Venindu-se după altora din celelalte conferințe, vor apărea în luna decembrie în revista META (www.metajournal.org).

title

Corneliu Bâlbă

Universitatea și spațiul public

text

Ivan Vuković

Rațiune și istorie: ideea kantiană de Universitate

text

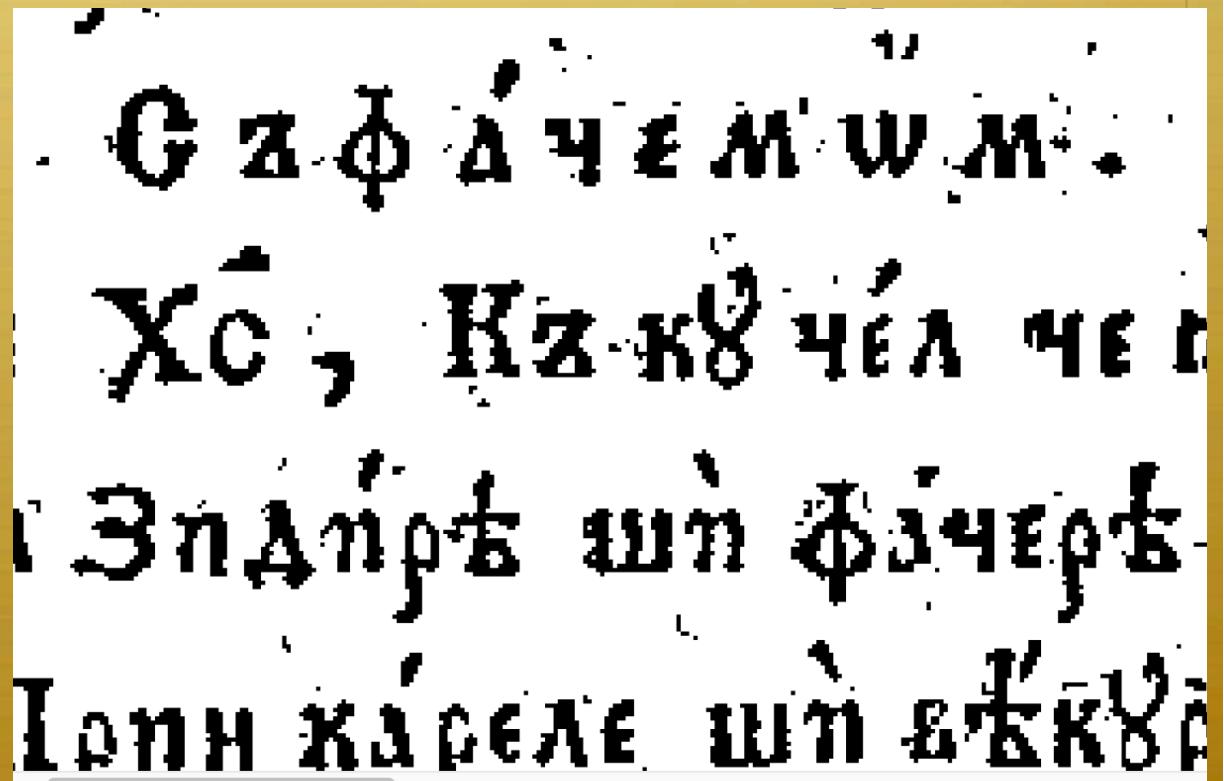
author

Pagini coordonate de George Bondor

```
<article>
  <title> Universitatea și spațiul public </title>
  <author> Corneliu Bâlbă </author>
  <text> ... </text>
</article>
<article>
  <title> Rațiune și istorie: ideea kantiană de Universitate </title>
  <author> Ivan Vuković </author>
  <text> ... </text>
</article>
```

P5. OCR on printed texts and manuscripts

- ◆ Target: decipher old Cyrillic-Romanian writings
- ◆ How?
 - ◆ segment
 - ◆ cluster
 - ◆ label
 - ◆ apply language models



Софія Мішні.
Хс, Кз-къчел че т
Ізлдирѣ шт фічерѣ
Іоти къреле шт вѣкъ

OCR - segmentation

С а ф д ү е м ү м

Хс, Қзжчел че
Зпдпрш шт фзчсрш
Іопн жарел€ шт в'кквр

OCR - clustering

Сәзбі ачылышын
Хс, Кз жөнчел үе т
Зілдірік шт **Ф**аңдері
Іорн жарелған шт 8'ккө

OCR - clustering

Салом! Менің аты менің
жөні, Каз-Жы^Чел Чесн
Зілдіркі шың фәс^Черк
Іорн жарелғаның екінші

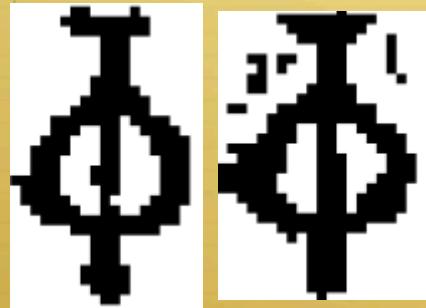
OCR - clustering

Сәғдәчелік
ХС, КЗЖЧЕЛ ЧЕ
ЗПДРК шт ФЗЧЕРК
Іорн жарелған екінші

OCR - clustering

Сәз-Фәдәүе-Мәшім.
Хс, **К**аз-**К**өчел ңе т
Зәлдірек шт Фәндер
Іспи **К**аседе шт 87**К**өж

OCR - labeling



⇒ F



⇒ t



⇒ r

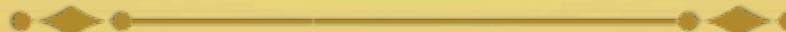


⇒ u

Features for clustering and labeling

- ◆ parameters in Cartesian coordinates (after norming) (https://en.wikipedia.org/wiki/Cartesian_coordinate_system)
 - ◆ rectangular area of the segment
 - ◆ ratio D_x/D_y
 - ◆ corners (upper left, upper right, lower left, lower right)
 - ◆ vertical stripes, horizontal stripes
 - ◆ occupied squared zones in the rectangular area
 - ◆ # of significant contiguous zones
 - ◆ position of weight centers of significant contiguous zones $\langle x,y \rangle$
 - ◆ ratio L_x/L_y of significant contiguous zones
 - ◆ position of weight center of the total black area
- ◆ same parameters in polar coordinates (after norming the radius) (https://en.wikipedia.org/wiki/Polar_coordinate_system)

Aplying language models



кътотъка · астъпъз · ръга лъвъ

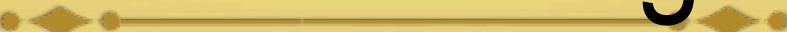


cu totul astupă guralor



cu totul astupă gura lor

P6. Transposing old Romanian into modern language

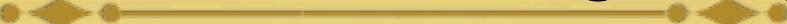


- ◆ Goal: bring an old Romanian text into a form that is easily read by a contemporary reader
- ◆ What does it mean?
 - ◆ reconstruct the old Romanian morphology
 - ◆ build a POS-tagger + lemmatizer for old Romanian
 - ◆ identify types of stylistic differences (?...)
- ◆ Biblio:
 - ◆ Radu Simionescu, Dan Cristea, Gabriela Haja (2012). Inferring diachronic morphology using the Romanian Thesaurus Dictionary - here: <https://profs.info.uaic.ro/~dcristea/papers/Inferring%20Diachronic%20Morpho-ConsILR%202011.pdf>

The Cyrillic-Romanian original

Челаде проповѣдь Знадіїї сѧе Ѥтѡарчємъ.
Чѣле маїкъ меснїре Прѣнсе Днї докитоаче.
Сѫнт маїмѹлт ижскатоаре. Пентрѹ ачѧ-
ста мѹлт ижскатоарѣ сѫнт Епѡрїї, шо жа-
преле чѣле салбатиче. шо ѿде чѣле сал-
батиче сѫнт ижскатоаре дѣ ѹбменї, ка сѧ
иѹлпѹбскъ иѣмѹлъ, келтѹндѹсъ дѣ ҳіа-
риле чѣле жижкатоаре дѣ кárne. Іјр чѣле
стриккатоаре ՚лтѡрадъ, сѫнт пѹцѹн ижскато-
аре. Днї кárе прѣчию лѧшанка дѣбїл сѫфа-
че маїкъ оѹнѹпъ леѹ сѹнгѹръ. Къ дѹз че къ

After transliteration After language update



Ci la privirea zidirii să ne întoarcem. **Ceale** mai cu lesnire prinse din dobitoace **sînt** mai mult născătoare. Pentru aceasta mult născători **sînt epurii**, și caprele **ceale** sălbatrice și oile **ceale** sălbatrice **sînt** născătoare de **geameni**, ca să nu lipsească neamul, cheltuindu-**să** de **hiarăle** **ceale** mîncătoare de carne. Iar **ceale** stricătoare altora, **sînt** puțin născătoare. Din care pricină leoaică de-abia **să** face maică unui leu singur.

Ci la privirea zidirii să ne întoarcem. **Cele** mai cu lesnire prinse din dobitoace **sînt** mai mult născătoare. Pentru aceasta mult născători **sînt iepurii**, și caprele **cele** sălbatrice și oile **cele** sălbatrice **sînt** născătoare de **gemeni**, ca să nu lipsească neamul, cheltuindu-**se** de **fiarele** **cele** mîncătoare de carne. Iar **cele** stricătoare altora, **sînt** puțin născătoare. Din care pricină leoaică de-abia **se** face maică unui leu singur.

Declination...

N-A, sg, n-art.: hiară
G-D, sg, n-art.: hiare

N-A, pl, n-art.: hiare
G-D, pl, n-art.: hiare

N-A, sg, art.: hiara
G-D, sg, art.: hiarei

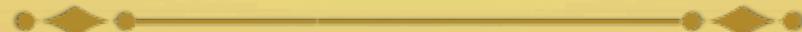
N-A, pl, art.: hiarăle
G-D, pl, art.: hiarălor

N-A, sg, n-art.: fiară
G-D, sg, n-art.: fiare

N-A, pl, n-art.: fiare
G-D, pl, n-art.: fiare

N-A, sg, art.: fiara
G-D, sg, art.: fiarei

N-A, pl, art.: fiarele
G-D, pl, art.: fiarelor



N-A, sg, n-art.: hiară
G-D, sg, n-art.: hiare

N-A, pl, n-art.: hiare
G-D, pl, n-art.: hiare

N-A, sg, art.: hiara
G-D, sg, art.: hiarei

N-A, pl, art.: hiarăle
G-D, pl, art.: hiarălor

N-A, sg, n-art.: fiară
G-D, sg, n-art.: fiare

N-A, pl, n-art.: fiare
G-D, pl, n-art.: fiare

N-A, sg, art.: fiara
G-D, sg, art.: fiarei

N-A, pl, art.: fiarele
G-D, pl, art.: fiarelor

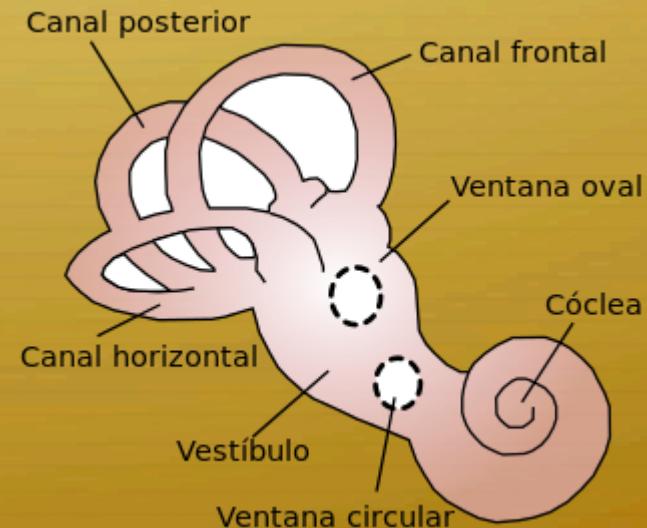
P7. Dialogue system on an ontology-based medical text



- ◆ Having a field ontology (general or specialist medicine) and a text that contains many specific references (medical terms), build a system able to carry out a dialogue with the user in natural language, the system being able to explain (define) terms from that text, identify references in and out of the text (provide links to other passages in text or sources of online information), to reformulate areas of the text in order to explain or summarize them.
- ◆ Possible extensions: The system carries the dialogue through a generated voice and a voice recognition application, the system can use text / resources in Romanian and English at the same time (automatic translation), eventually the system can generate and evaluate a test originating in a text.

Example of a possible interaction

- ◆ BOT: What is the inner ear?
- ◆ The inner part of the ear, I do not know more details.
- ◆ BOT: You still know something about the cochlea?
- ◆ No, please tell me more about the cochlea.
- ◆ BOT: Cochlea or membranous snail is a component part of the inner ear, made up of a spiral grooved groove that gives it a snail shape. Cochlea collects hearing information, so it is the place where the hearing receptor is located.



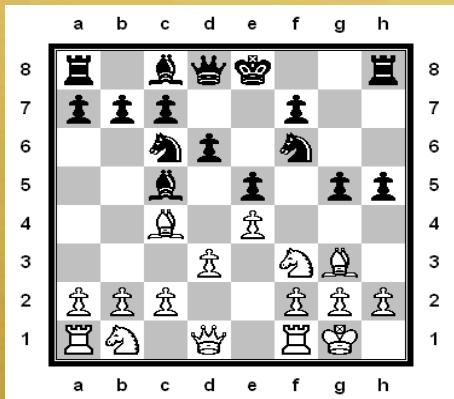
P8. The chess game



- ◆ A system capable of evaluating situations during a chess game and providing suggestions on optimal moves, indicating possible mistakes, generating a comment in natural language on a game.
- ◆ Eventually, the system includes several chess strategies and indicates the moves suggested by each strategy, explaining the decision made.
- ◆ Possible extensions: a mobile device interface that allows the system to be used as an assistant in a real-life chess game (with or without the automatic recognition of the position of the pieces on the board)

Examples of possible interactions

- ◆ User defines a configuration, for instance:



- The system suggests that the Hf3xg5 is the optimal move
- The system explains the suggestion according to MINMAX, AlfaBeta, Negamax, others. The system can indicate the precision of the suggestion, the number of anticipated moves, the following suggested moves.

- ◆ User introduces a game:

1.e4 e5 2.Nf3 f6 3.Nxe5 fxe5 4.Qh5+ Ke7 5.Qxe5+ Kf7 6.Bc4+ d5 7.Bxd5+ Kg6 8.h4 h5 9.Bxb7 Bxb7 10.Qf5+ Kh6 11.d4+ g5 12.Qf7 Qe7 13.hxg5+ Qxg5 14.Rxh5# 1-0

- ◆ System comments / explains the game indicating the optimal moves, mistakes, situations where other moves would lead to victory.

P9. Voice-text aligning algorithms



- ◆ System capable of aligning a voice recording with the corresponding text. The system must identify areas in the audio or text that do not correspond to the parallel resource or areas that contain content that does not correspond to the language known to the system (nonspecific phoneme, pronunciation errors, noises).
- ◆ Extensions: The system allows indexing and compacting aligned resources by identifying "perfect" identities between voice-text aligned pairs. Eventually, the system also lets you generate a text-based on a voice signal or recognize a voice signal.
- ◆ The method must be independent of the speaker.

Alignment example

