

Kaip paruošti savo duomenis automatiniam vertimui?

Andrejs Vasiljevs, ELRC/Tilde

Bendraautoriai:

Khalid Choukri, ELRA

Josef van Genabith, DFKI

An iceberg floating in the ocean. The small tip above the water represents the visible web (4%), while the much larger mass below the surface represents the deep web (96%).

4%

Viešasis
žiniatinklis

96%

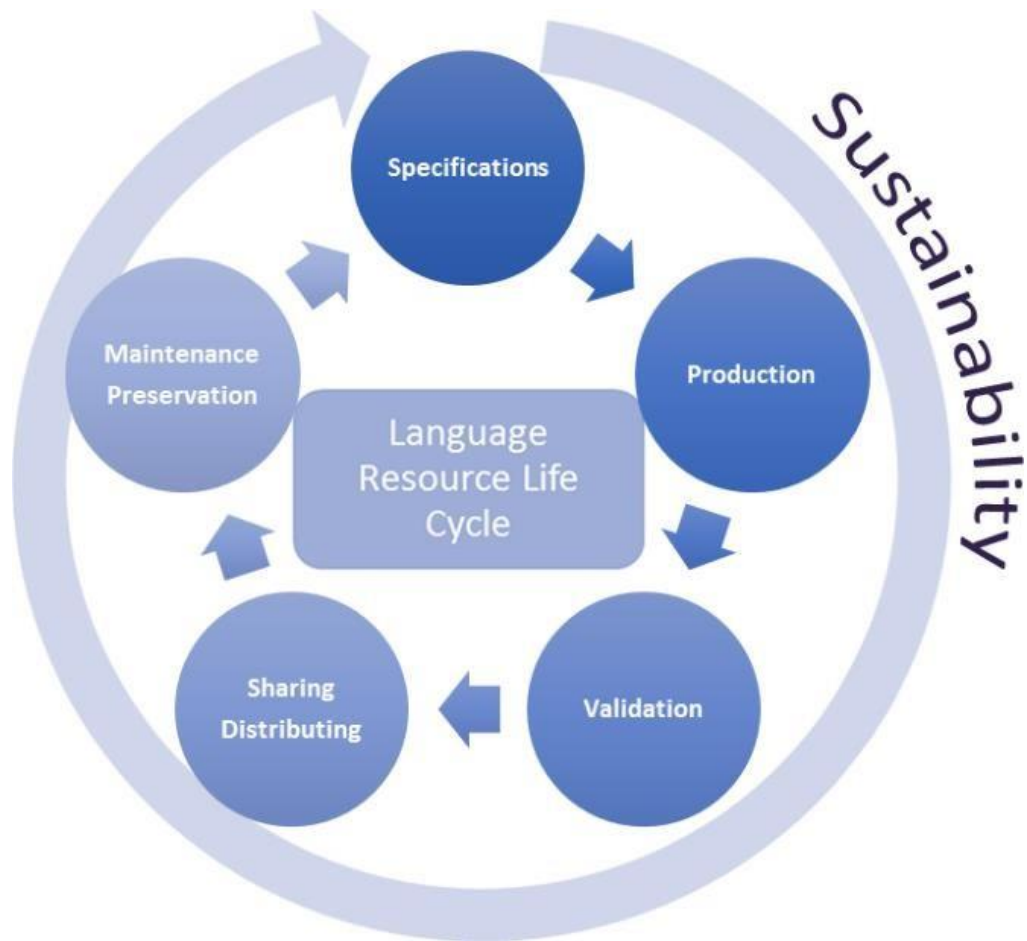
Gilasis
žiniatinklis

Daugiakalbės duomenų bazės
Viešojo sektoriaus ištekliai
Organizacijų saugyklos
Teisiniai dokumentai
Mokslinės ataskaitos
Ambulatorinės kortelės
Ir kt.

*Informacija, esanti institucijų
viduje arba internete,
apsaugota slaptažodžiu*

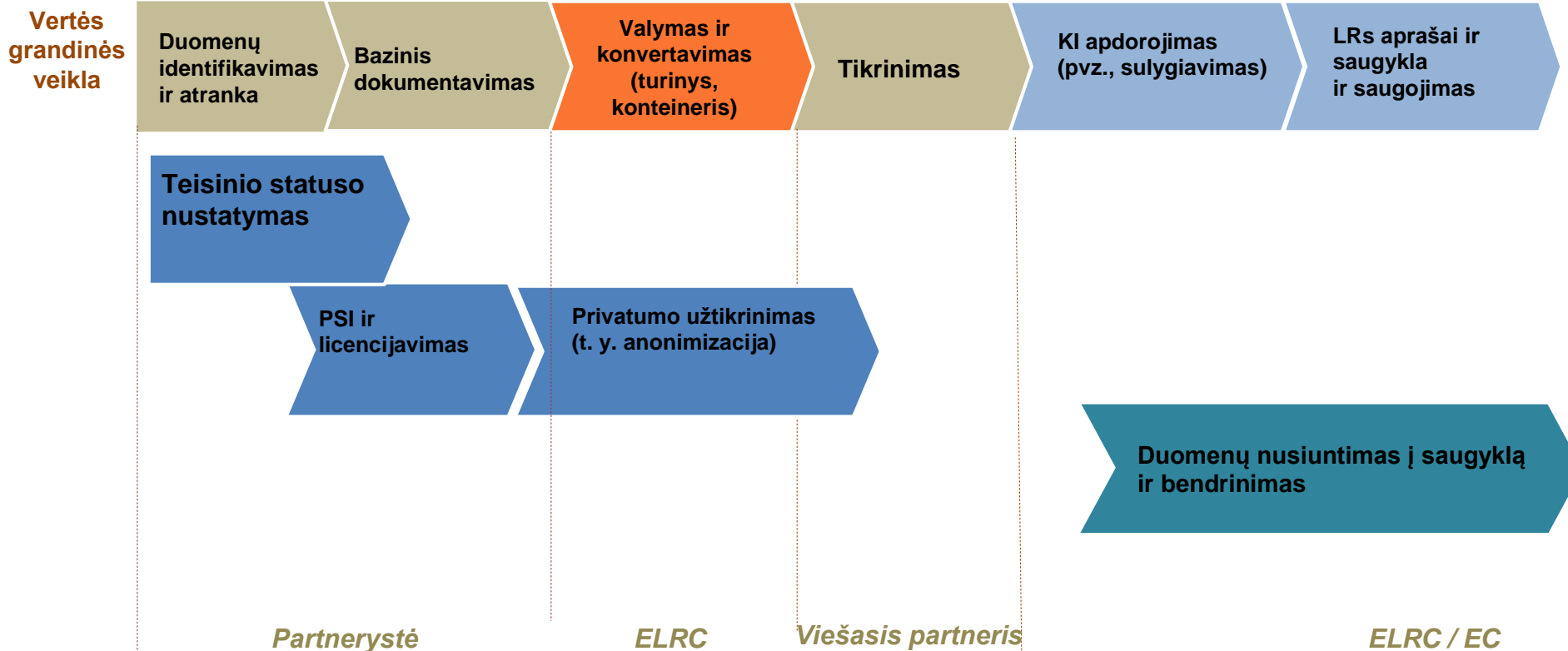


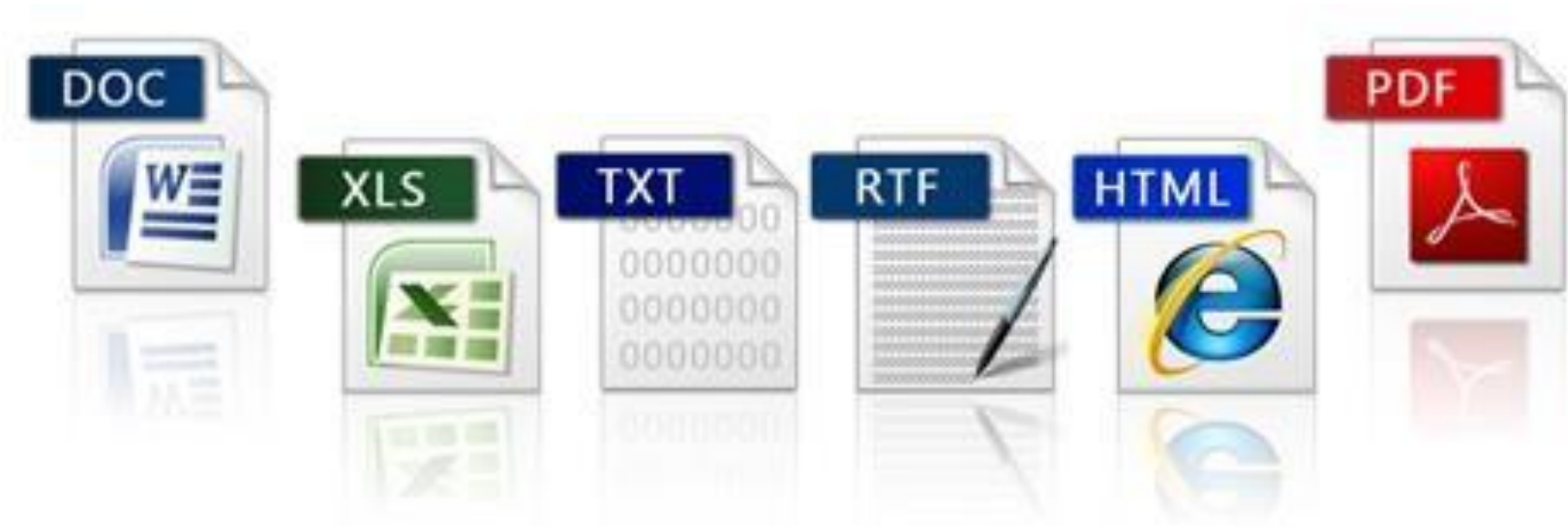
- Matomi duomenys, pvz., [žiniatinklis](#)
 - viešosios svetainės
 - lankstinukai, brošiūros, naujienos ir kt.
- [Nematomi duomenys](#): archyvai, paslėptas žiniatinklis, vidinės saugyklos
 - išversti dokumentai, ataskaitos, kalbos, susitikimų protokolai ir kt.
- Duomenys, gauti iš vertimų, užsakytų išoriniams teikėjams:
 - iš [kalbinių paslaugų teikėjų](#) ir laisvai samdomų vertėjų
 - vertimo atmintys, pateikiamos pagal sutartį
 - atitinkamos sąlygos sutartyse dėl iš išorės užsakytų vertimų



Duomenų pakavimo darbų eiga

Duomenys → KI (kalbiniai ištekliai)

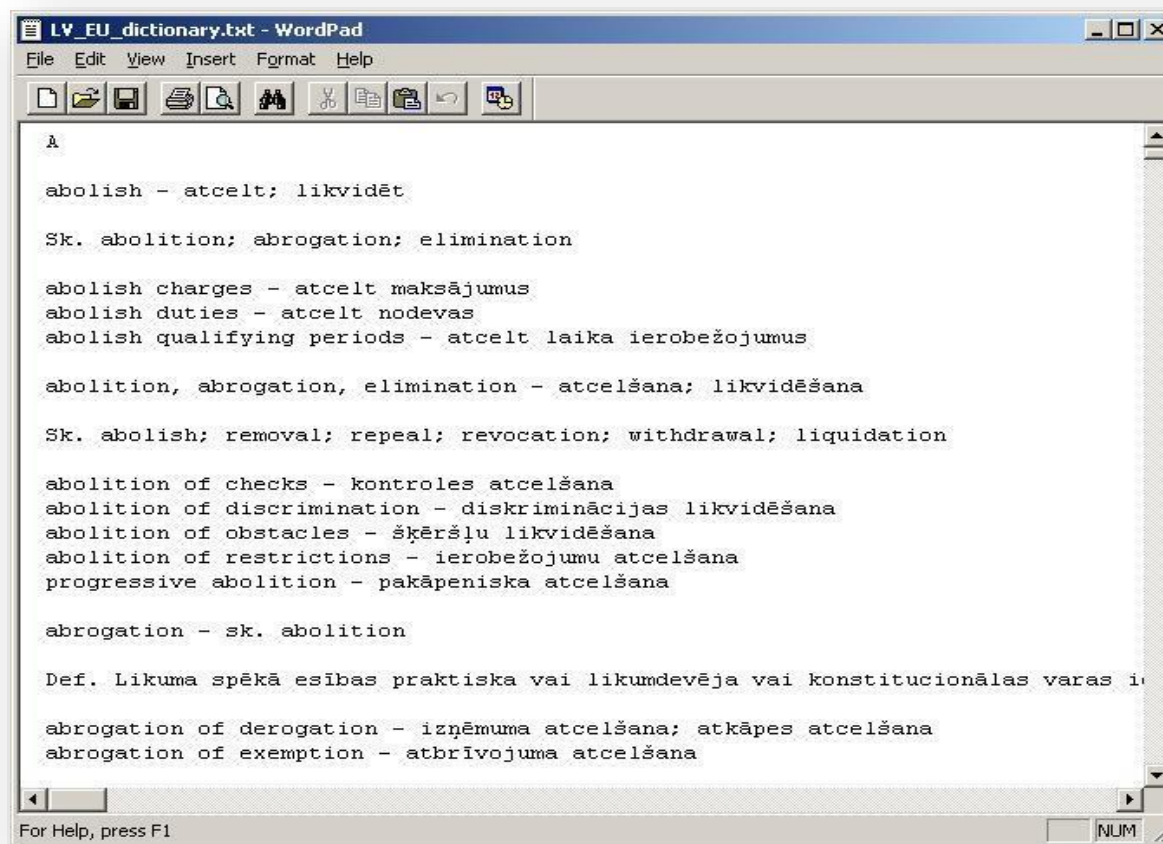


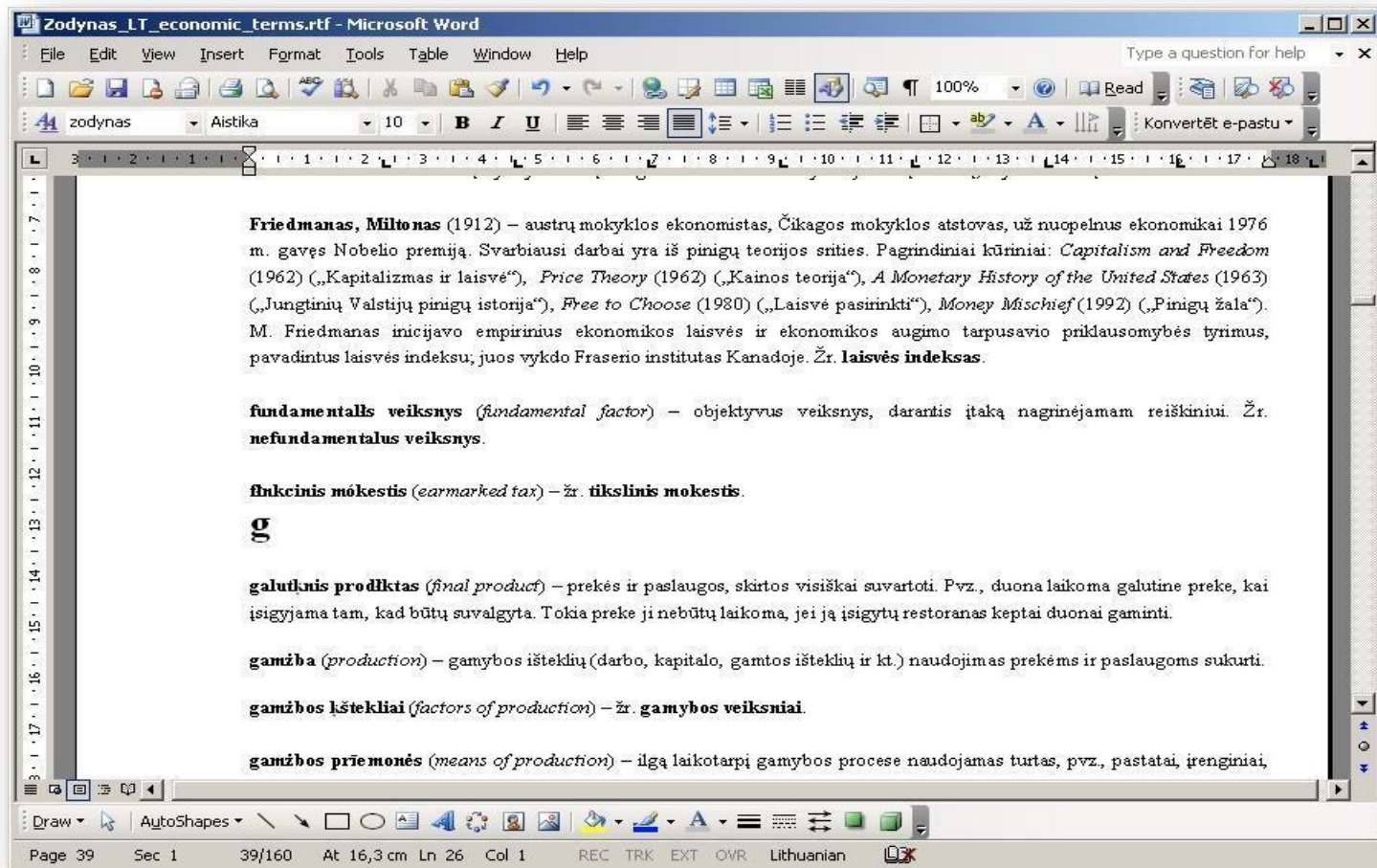


**Skaitmeniniai tekstiniai duomenys įvairiais
formatais – vertingi kalbiniai
ištekliai**

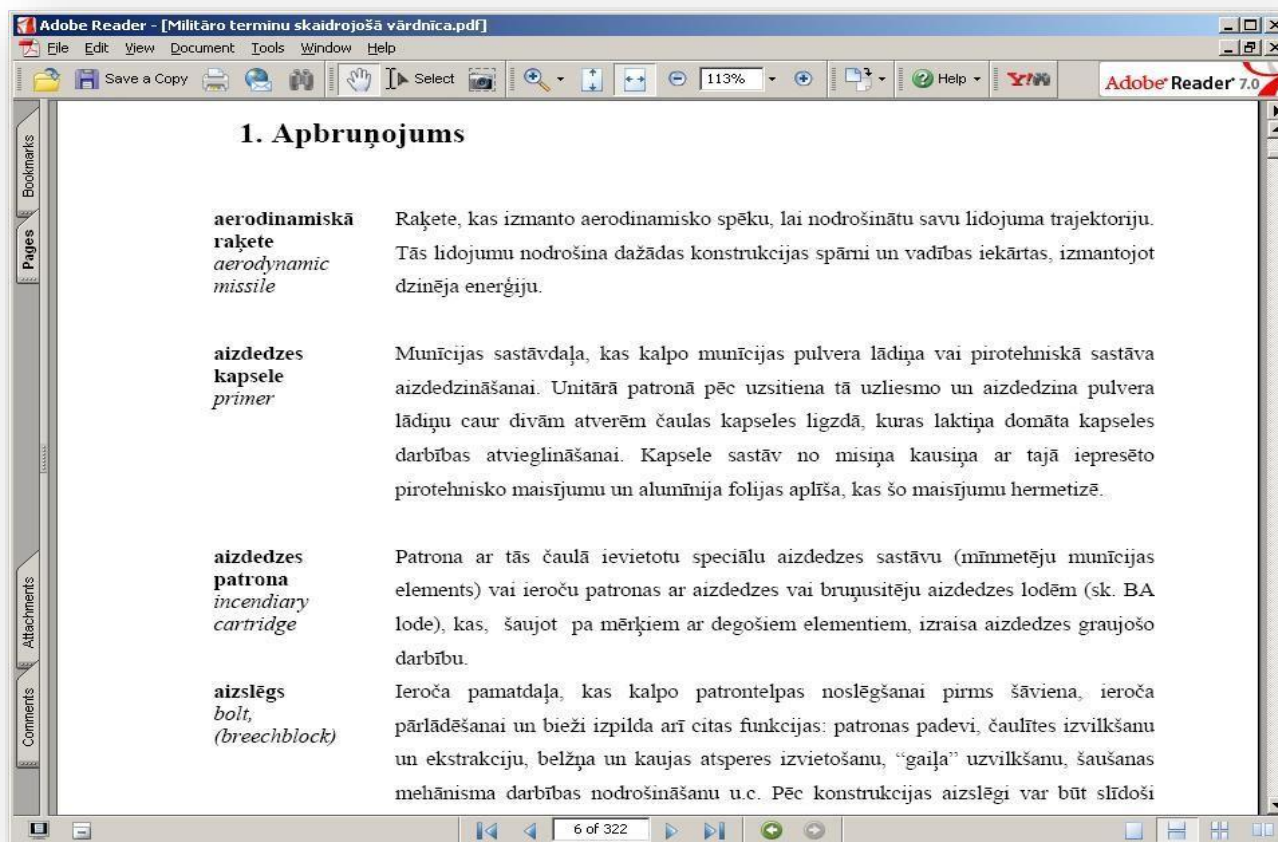
EE-EN-RU matematiniai terminai spausdintine forma

L 361 lāhenema	106	L 388 lūliti	M1 maagiline	107	M 33 majoreerimine
L 361 lāhenema piirvārtusele	approach a limit	приближаться к пределу	M		
L 362 lāhenemine	approaching	приближение	M 1 maagiline ruut	magic square	магический квадрат
L 363 lāhim	closest, nearest	ближайший	M 2 maastikuformaat	landscape size	горизонтальный формат
L 364 lāhis-	adjacent	прилежащий	M 3 maatriks	matrix	матрица
L 365 lāhis- (= lāhend-)	approximate	приближенный	M 4 maatriks-	matrix	матричный
L 366 lāhiskaaket	adjacent side	прилежащий катет	M 5 maatriksalgebra	matrix algebra	матричная алгебра
L 367 lāhiskūlg	adjacent side	прилежащая сторона	M 6 maatriksesitus	matrix representation	матричное представление
L 368 lāhismurd	convergent	подходящая дробь	M 7 maatriksi astak	rank of a matrix	ранг матрицы
L 369 lāhte-	initial, original, starting	исходный, начальный	M 8 maatriksi diagonaal	diagonal of a matrix	диагональ матрицы
L 370 lāhteandmed (⇒ algandmed)	initial data, source data	исходные данные, начальные данные	M 9 maatriksi elementaarteisendused	elementary operations on matrices	элементарные операции над матрицами
L 371 lāhtehulk	domain, initial set	множество отправления	M 10 maatriksi jalg	spur of a matrix, trace of a matrix	след матрицы
L 372 lāhtepunkt (= alguspunkt)	initial point, starting point	исходная точка, начальная точка, отправная точка	M 11 maatriksi pōoramine	inversion of a matrix, matrix inversion	обращение матрицы
L 373 lāhtesimbol	initial symbol	начальный символ	M 12 maatriksite liitmine	addition of matrices	сложение матриц
L 374 lāhtevōrand	original equation	исходное уравнение, первоначальное уравнение	M 13 maatriksite ring	ring of matrices	кольцо матриц
L 375 lāhtuma	start	исходить	M 14 maatriksmāng	matrix game	матричная игра
L 376 lāvi	threshold	порог	M 15 maatrikstāhistus	matrix notation	матричная символика
L 377 lūhem telg (⇒ vākkelgelg)	minor axis	малая ось	M 16 maatriksvārtustega funktsioon	matrix-valued function	матрично-значная функция
L 378 lūhend	abbreviation	аббревиатура, сокращение	M 17 Mackey topoloogia	Mackey topology	топология Макки
L 379 lūhendama	abbreviate, shorten	сокращать, укорачивать	M 18 Maclaurini rida	Maclaurin's series	ряд Маклорена
L 380 lūhendamine	abbreviation, shortening	сокращение, укорачивание	M 19 madalaim numbrikoht	least significant digit	младший разряд
L 381 lūhendatud	abbreviated, abridged, contracted	сокращенный	M 20 madalamat jārku	of lower order	нижнего порядка
L 382 lūhendatud korrutamine	abbreviated multiplication	сокращенное умножение	M 21 magasin (= pinu)	stack	магазин, стек
L 383 lūhendatud tāhis	abridged notation, contracted notation	сокращенное обозначение	M 22 magasin algoritim	stack algorithm	магазинный алгоритм
L 384 lūhike	short	короткий	M 23 magasin automaat	push-down automaton	магазинный автомат
L 385 lūhim	shortest	кратчайший	M 24 magasin tipp	top of stack	вершина стека
L 386 lūke (⇔ translatsioon)	shift, slide, translation	перемещение, перенос, сдвиг	M 25 magistraal	turnpike	магистраль
L 387 lūli	link	звено	M 26 maaha tōmbama	cross out	вычеркивать
L 388 lūliti	switch	переключатель	M 27 maahatōmbamine	crossing out	вычеркивание
			M 28 maht (= mahukas)	capacity	ёмкость
			M 29 mahukas	capacious	ёмкий
			M 30 mahukesmine	power mean	степенное среднее
			M 31 majorant	majorant	мажоранта
			M 32 majoreerima	majorize	мажорировать
			M 33 majoreerimine	majorizing	мажорирование





LV karinē terminoloģija PDF formatu



Adobe Reader - [Militāro terminu skaidrojošā vārdnīca.pdf]

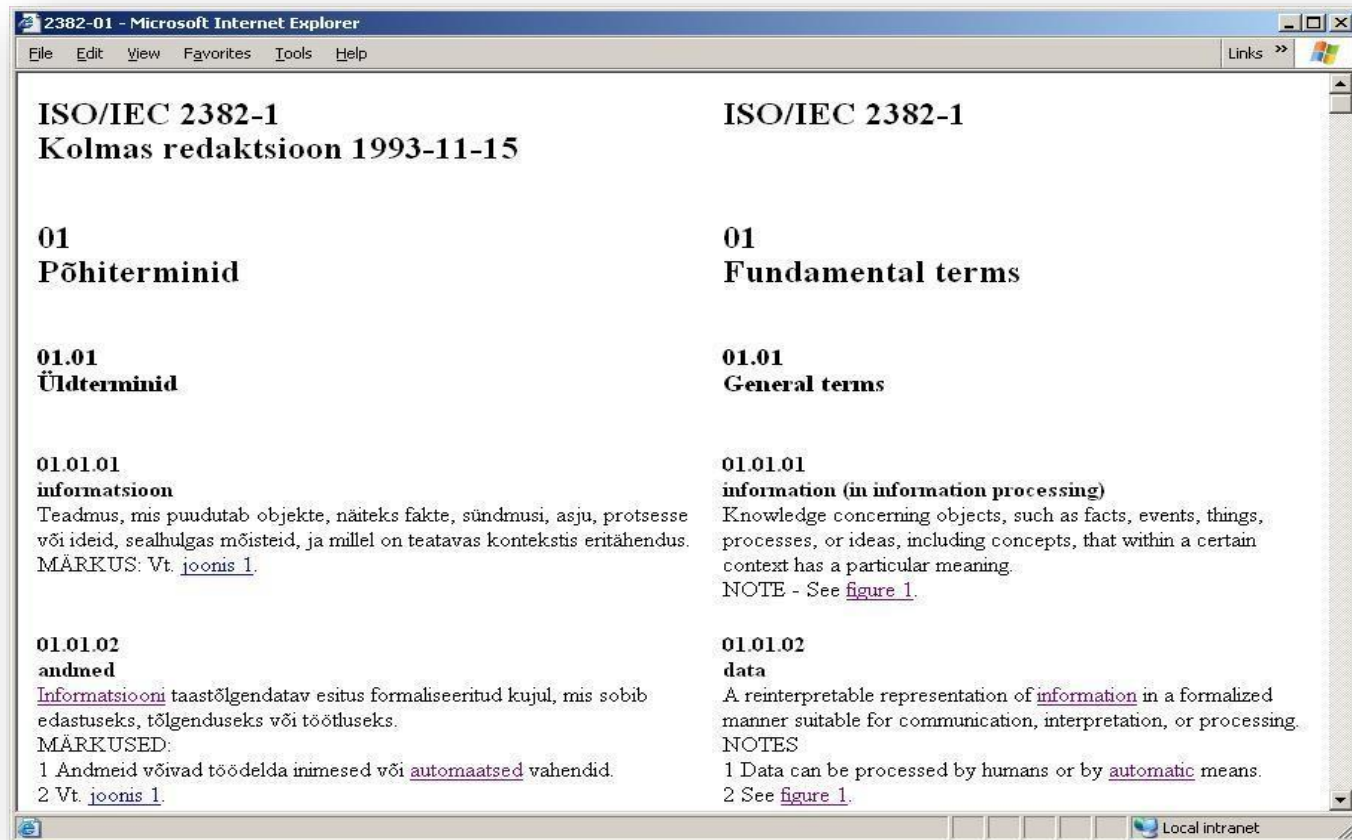
File Edit View Document Tools Window Help

Save a Copy Select 113% Help Adobe Reader 7.0

1. Apbruņojums

aerodinamiskā raķete <i>aerodynamic missile</i>	Raķete, kas izmanto aerodinamisko spēku, lai nodrošinātu savu lidojuma trajektoriju. Tās lidojumu nodrošina dažādas konstrukcijas spāni un vadības iekārtas, izmantojot dzinēja enerģiju.
aizdedzes kapsle <i>primer</i>	Munīcijas sastāvdaļa, kas kalpo munīcijas pulvera lādiņa vai pirotehniskā sastāva aizdedzināšanai. Unitārā patronā pēc uzsietiena tā uzliesmo un aizdedzina pulvera lādiņu caur divām atverēm čaulas kapseles ligzdā, kuras laktiņa domāta kapseles darbības atvieglināšanai. Kapsule sastāv no misiņa kausiņa ar tajā iepresēto pirotehnisko maisījumu un alumīnija folijas aplīša, kas šo maisījumu hermetizē.
aizdedzes patrona <i>incendiary cartridge</i>	Patrona ar tās čaulā ievietotu speciālu aizdedzes sastāvu (mūnmetēju munīcijas elements) vai ieroču patronas ar aizdedzes vai bruņusitēju aizdedzes lodēm (sk. BA lode), kas, šaujot pa mērķiem ar degošiem elementiem, izraisa aizdedzes graujošo darbību.
aizslēgs <i>bolt, (breachblock)</i>	Ieroča pamatdaļa, kas kalpo patronatelpas noslēgšanai pirms šāviena, ieroča pārlādēšanai un bieži izpilda arī citas funkcijas: patronas padevi, čaulītes izvilkšanu un ekstrakciju, belzņa un kaujas atsperes izvietošanu, "gaiļa" uzvilkšanu, šaušanas mehānisma darbības nodrošināšanu u.c. Pēc konstrukcijas aizslēgi var būt slidoši

6 of 322



2382-01 - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links »

ISO/IEC 2382-1 Kolmas redaktsioon 1993-11-15	ISO/IEC 2382-1
01 Põhiterminid	01 Fundamental terms
01.01 Üldterminid	01.01 General terms
01.01.01 informatsioon Teadmus, mis puudutab objekte, näiteks fakte, sündmusi, asju, protsesse või ideid, sealhulgas mõisteid, ja millel on teatavas kontekstis eritähendus. MÄRKUS: Vt. joonis 1 .	01.01.01 information (in information processing) Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning. NOTE - See figure 1 .
01.01.02 andmed Informatsioon taastõlgendatav esitus formaliseeritud kujul, mis sobib edastuseks, tõlgenduseks või töötamiseks. MÄRKUSED: 1 Andmeid võivad töödelda inimesed või automaatsed vahendid. 2 Vt. joonis 1 .	01.01.02 data A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. NOTES 1 Data can be processed by humans or by automatic means. 2 See figure 1 .

Local intranet

LV-RU įvairių domenų terminų sąrašas „Excel“ formatu



Microsoft Excel - Dainu skapis 2006 Jan.xls

File Edit View Insert Format Tools Data Window Help

Type a question for help

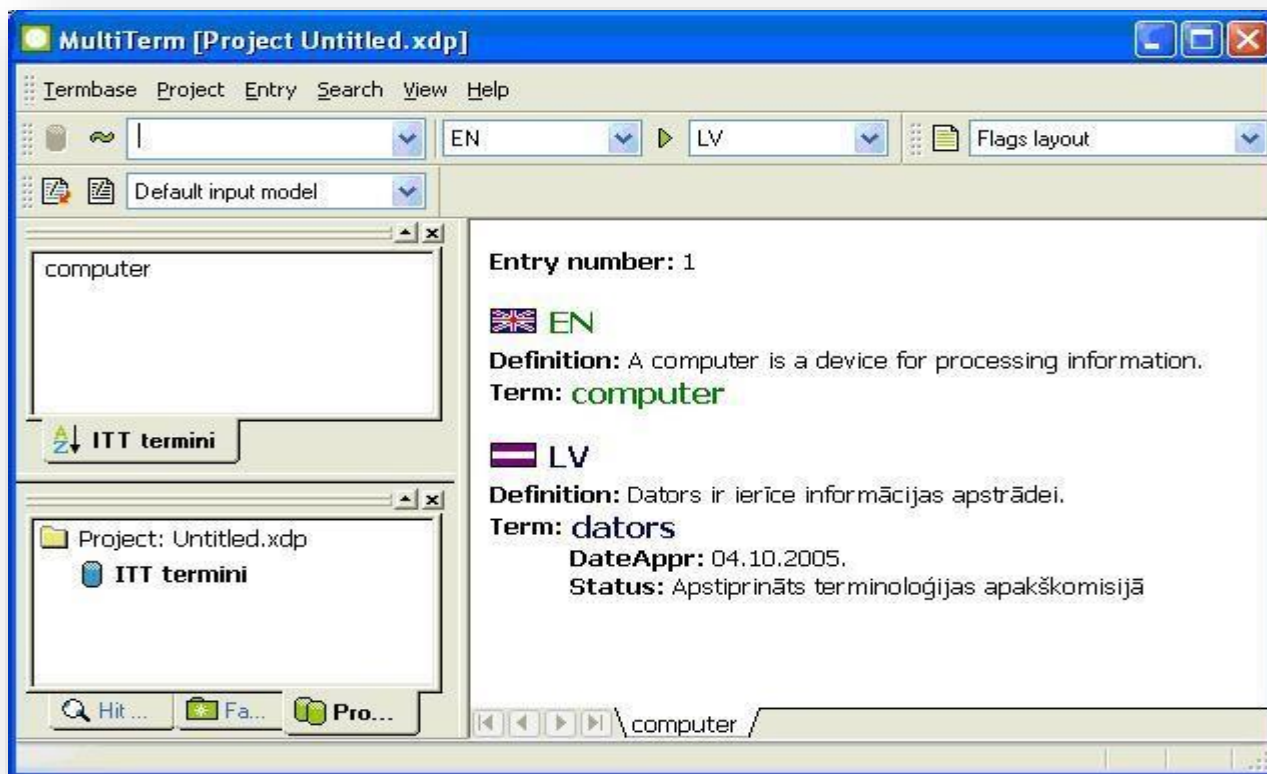
Arial 8 B I U

A8 diagnosticēšanas noteikumi

	A	B	C	D	E	F	G
977	elektronoptiskais pārveidotājs	электронно-оптический преобразователь	krimināl.	1979			10
978	hidropārveidotājs	гидропреобразователь	lauks. tehn.	1974			33
979	induktīvais pārveidotājs	электродинамический преобразователь	elektron.	1978			26
980	induktīvais pārveidotājs	индуктивный преобразователь	elektron.	1980			40
981	kapacitīvais pārveidotājs	емкостной гидропреобразователь	elektron.	1980			40
982	kapacitīvais pārveidotājs	электростатический преобразователь	elektron.	1978			26
983	magnetrokšņveidojošais pārveidotājs	магнитострикционный гидропреобразователь	elektron.	1978			14
984	neapgriežamais pārveidotājs	необратимый преобразователь	elektron.	1978			26
985	pjezoelektriskais pārveidotājs	пьезоэлектрический преобразователь	elektron.	1978			14
986	pneimohidropārveidotājs	пневмогидропреобразователь	lauks. tehn.	1974			33
987	signālu pārveidotājs	устройство преобразования сигналов; УПС	elektron.	1978			18
988	pārvēlums	перекат	šūš.	1980			20
989	kursa pārvešana	перевод румбов	jūr.	1981			8
990	meridionālais siltuma pārvietojošs	межширотный перенос	meteor.		38	120	
991	ārpusrises pārvietošanās trīsdimensiju telpā	внетрассовое перемещение в трехмерном пространстве	ek. ģ.	1975			14
992	atpakāļējā pārvietošanās	"попятное" перемещение	ek. ģ.	1975			14
993	ciklona anomālā pārvietošanās	анормальное перемещение циклона	meteor.		38	120	
994	kravas pārvietošanās	смещение груза	jūr.	1981			9
995	pierobežas pārvietošanās	пограничное перемещение	ek. ģ.	1975			14
996	sadalošā pārvietošanās	распределительное перемещение	ek. ģ.	1975			14
997	laikapstākļu pārvietošanās; laikapstākļu pārnese	перенос погоды	meteor.		38	120	
998	mitruma pārvietošanās atmosfērā	перенос влаги в атмосфере	meteor.		38	120	

Ready NUM


```
DH-KOZGE.SGH - Notepad
File Edit Format View Help
<dh-kozge>
<entry>
<orth>Abbuchungsauftrag</orth>
<sense>Lehkvjsi megbkzjs. A szjmlavezeto banknak adott visszavonhato megbkzjs, meghatjrozott osszegnek a folyoszjmljról való lehkvjsjra.</sense>
</entry>
<entry>
<orth>Aberdepot</orth>
<orthvar>
<orth>Summenverwahrung</orth>
</orthvar>
<sense>Értékpapkrök bank jlteni megörzésének a gyakorlatban kevésbé alkalmazott módja, amikor az értékpapkr betevö ügyfél nem ugyanannak az értékpapkrnak, hanem csak ugyanolyan minőségü és mennyiségü értékpapkrnak visszavételére jogosult. Nincs szó tehát igazi letétről, a betett értékpapkr tulajdonjoga a megörző banké lesz.</sense>
</entry>
<entry>
<orth>Abfallbeseitigung</orth>
<sense>Hulladék eltjvolktjsa. Hjtartjsi, ipari és mezögazdasjgi hulladék begyüjtése, kezelése és tjroljsa. Az NSZK-ban a tartomjnyok kötelesek terveket készkteni a hulladék eltjvolktjsjra, feltüntetve a hulladéktjrolók helyét is.</sense>
</entry>
<entry>
<orth>Abfertigung</orth>
<sense>Irukezelés. A szjllktjsban az jru felvétele, jogilag a fuvarozjsi szerződés előkészktése és megkötése.</sense>
</entry>
<entry>
<orth>Abfertigungsgebühr</orth>
<sense>kezelési költség. A fuvarozö jlteni a feladjsi és a rendeltetési jllojmon végzett szolgjltatjsok ellenértéke. A kezelés és a szjllktjs költsége képezi a fuvardkjat.</sense>
</entry>
<entry>
```





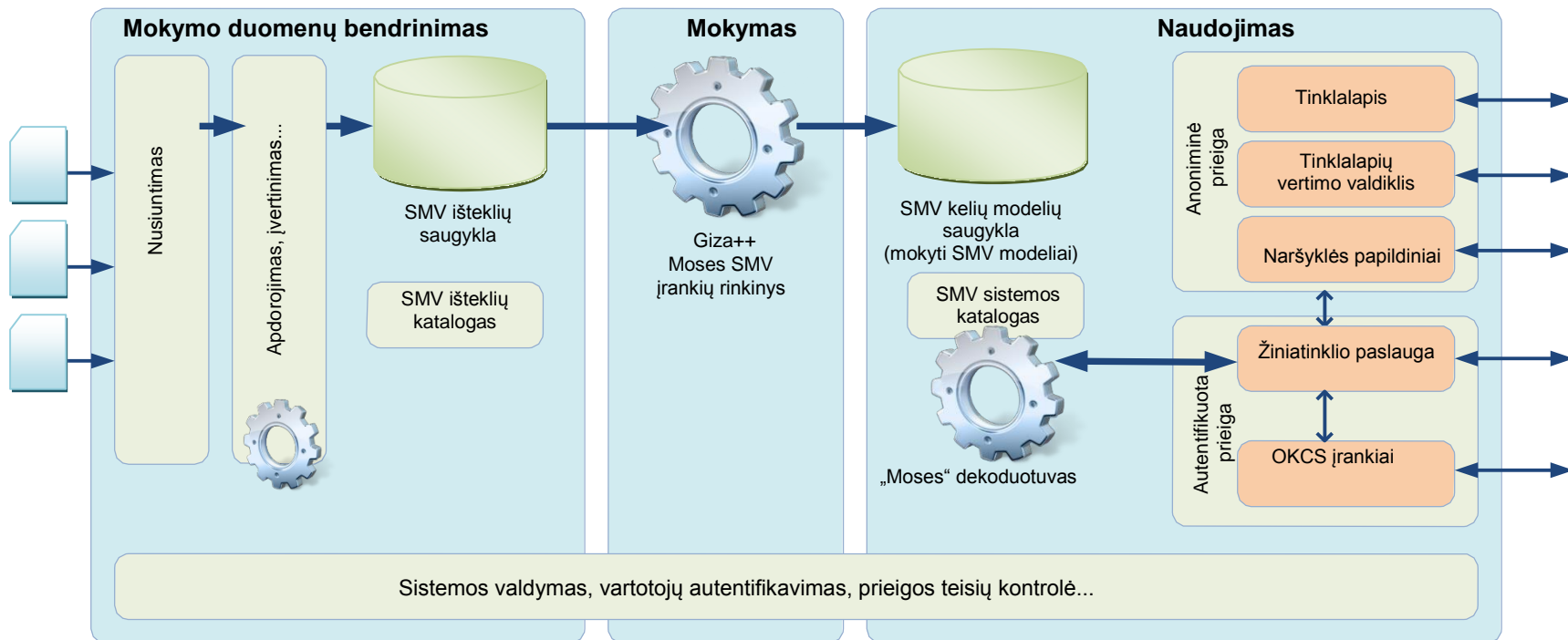
- Šaltinių identifikavimas, duomenų rinkinių (neapdorotų duomenų) identifikavimas ir atranka
 - duomenis galima gauti iš matomų šaltinių
pvz., išgauti iš žiniatinklio
 - duomenis gali pateikti viešojo sektoriaus institucijos
 - viešojo sektoriaus institucijos gali padidinti matomų išteklių identifikavimą
- Gali būti atliekama bendradarbiaujant ELRC ir duomenų teikėjui

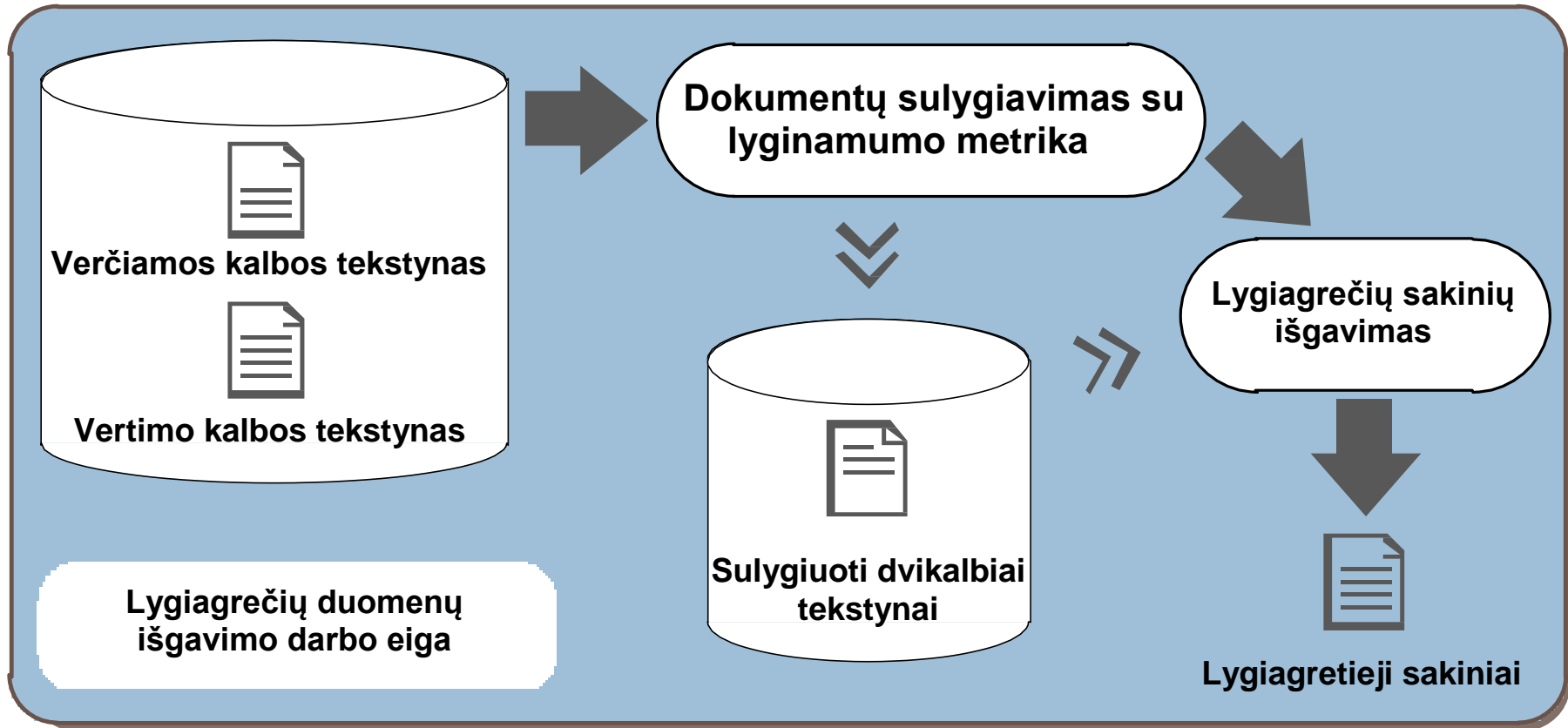
- Duomenų formato **išvalymas**
Formatavimo atsisakymas, simbolių rinkinių kodavimas, pvz., UTF8, formatavimo priemonės, pvz., paryškintasis, pasvirasis šriftas; grafika, skelbimai, lentelės, HTML žymos ir kt.
- Failų **konvertavimas**
pvz., konvertavimas į XML, XLIFF ir kt.
- **Duomenų parengimas** automatinio vertimo įrankiams
pvz., išgavimas ir sulygiavimas
- Išvesties **tikrinimas** ir kokybės kontrolė
Kalbos išteklių formatas, turinys, saugojimas
- Kalbos išteklių **aprašas** (metaduomenys)
- Pakavimas ir **pristatymas** (duomenų saugykla su el. bendrinimu) EK ir savininkui

„LetsMT“ technologija



Let's MT!





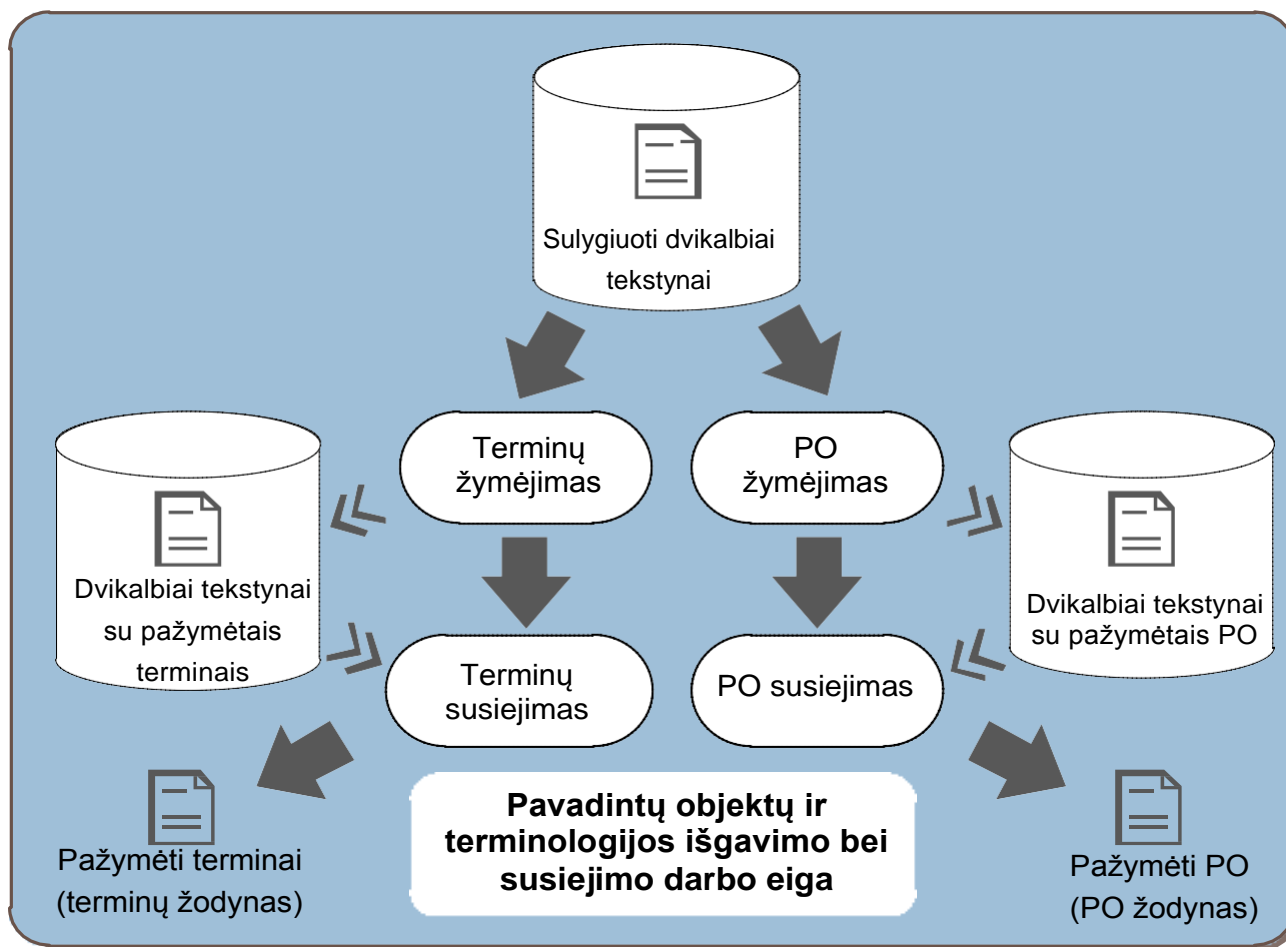
Lyginamųjų tekstynų apdorojimas norint išgauti lygiagrečiųjų sakinių poras

Sakinių iš lygiagrečiųjų dokumentų sulygiavimas



en	el
Greece of art and science	Η Ελλάδα των τεχνών και της επιστήμης
Greece is a place of culture, the arts and sciences.	Η Ελλάδα αποτελεί έναν χώρο πολιτισμού, τέχνης και επιστημών.
Its tradition of contribution to global cultural and scientific communities, combined with its outstanding natural beauty and excellent infrastructure, has made it an ideal place in which to hold conferences.	Η μακράιωνη συμβολή της στο παγκόσμιο γίνεσθαι, σε συνδυασμό με το μοναδικό φυσικό κάλλος και τις άριστες υποδομές, την καθιστούν ιδανικό τόπο διεξαγωγής συνεδρίων.
Over the last few years, Greece has more and more frequently welcomed people of letters, sciences and the arts, who have participated in symposia, conferences and exhibitions.	Τα τελευταία χρόνια, η ελληνική επικράτεια υποδέχεται όλο και συχνότερα ανθρώπους των γραμμάτων, των επιστημών και των τεχνών, οι οποίοι συμμετέχουν σε συμπόσια, συνέδρια και εκθέσεις.
Athens International Airport 'Eleftherios Venizelos', one of the most modern airports in the world in operation since 2001, greatly boosted the organization of international conferences.	Ο Διεθνής Αερολιμένας Αθηνών «Ελευθέριος Βενιζέλος», ένα από τα πλέον σύγχρονα αεροδρόμια παγκοσμίως, ο οποίος λειτουργεί από το 2001, έδωσε μεγάλη ώθηση στη διοργάνωση διεθνών συνεδρίων.
Conference tourism is extremely interdependent: it requires of course a high level of background support from the host country, and at the same time it can actively contribute to improving the overall standard of services in the region.	Ο συνεδριακός τουρισμός είναι άκρως αλληλεπιδραστικός: απαιτεί, βέβαια, ένα υψηλού επιπέδου υπόβαθρο από τη χώρα υποδοχής, ταυτόχρονα όμως συμβάλλει ενεργά στην αναβάθμιση της συνολικής ποιότητας μιας περιοχής.
It is logical that a country chosen as a conference location should be involved in the cultural 'product', giving the public, both residents and visitors, the chance to experience human achievement and innovative thought.	Είναι λογικό, ένας χώρος ο οποίος προτιμάται για τη διεξαγωγή συνεδρίων, να μετέχει προνομιακά στο πολιτιστικό «προϊόν», μιας και δίνει τη δυνατότητα σε κοινό, κατοίκους και επισκέπτες, να έρθουν σε επαφή με τα ανθρώπινα επιτεύγματα και τις καινοτομίες.
The Greece of the pre-Socratic philosophers, of the great poets, of Pheidias the sculptor and Asclepius the physician, extends its hospitality and its warmest welcome, honouring people of intellect and creativity, commerce and scientific progress.	Η Ελλάδα των προσωκρατικών φιλοσόφων, των μεγάλων ποιητών, του Φειδία και του Ασκληπιού υποδέχεται φιλόξενα και τιμά τους ανθρώπους του πνεύματος, του εμπορίου και της προόδου.
Scientific conferences in the land that gave birth to science	Συνέδρια στη χώρα που γέννησε τις επιστήμες
Greece has a large number of esteemed scientists, both here in the country and abroad.	Η Ελλάδα διαθέτει μεγάλο και υψηλής αξίας επιστημονικό δυναμικό, τόσο εντός όσο και εκτός συνόρων.
Greek scientists, with their inventions, innovations and research work, play a leading part in the international scientific community.	Οι Έλληνες επιστήμονες, με τις εφευρέσεις, τις καινοτομίες και το ερευνητικό τους έργο πρωταγωνιστούν στη διεθνή επιστημονική κοινότητα.
Numerous important scientific conferences take place in Greece, reflecting the significance the country places on innovative science.	Τα επιστημονικά συνέδρια που λαμβάνουν χώρα στην Ελλάδα είναι και πολλά και σημαντικά, αντανakλώντας τη σημασία που δίνει η χώρα στις καινοτόμες επιστήμες.
Medical, architectural, natural and humanistic scientific conferences enrich Greece's cultural life, and at the same time give participants the opportunity to experience the	Ιατρικά συνέδρια, αρχιτεκτονικά, φυσικών και ανθρωπιστικών επιστημών, πλουτίζουν την πολιτιστική ζωή της

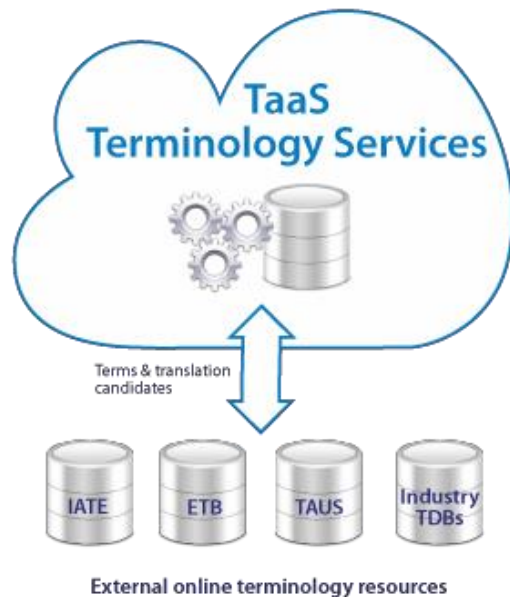
ACCURAT įrankiai pavadintiems objektams (PO) ir terminologijai išgauti



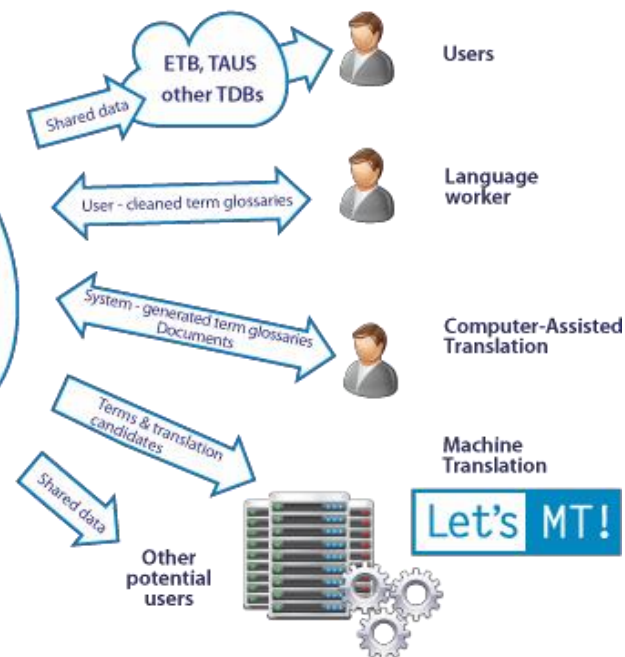
ACQUIRE & PROCESS



PROCESS, STORE & SHARE



REUSE





- Teisiniai klausimai
 - Teisinio statuso nustatymas (sprendimas priimti / atmesti)
 - Tikslumas ir privatumo užtikrinimas (pvz., anonimizacija)
 - PSI taikymas kaip licencijų alternatyva
- ELRC konsorciumo vaidmuo
 - Pagalba praktiniais klausimais
 - Techninė / teisinė pagalbos tarnyba, konsultavimas
 - Licencijų sutarčių pavyzdžiai
 - Valstybinių institucijų atvirosios licencijos
 - Standartinės pakartotinio naudojimo licencijos
 - Licencijų tarpusavio suderinamumas



- Nustatyti didelį duomenų apie asmenis, organizacijas ir kt. šaltinį
- Naudoti pavadintų objektų atpažinimo priemonę norint rasti ir pašalinti privačius biografinius duomenis (vardus, vietas, datas, gimimo informaciją ir kt.) ir juos pakeisti bendraisiais vietos rezervavimo ženklais.
- Patvirtinti, kad rezultatai atitinka priimtinus reikalavimus
 - atmesti duomenis, jei anonimizacija nepakankamai tiksli

– Bendrinimas / platinimas

- Įsitikinkite, kad jūsų duomenys atitinka PSI direktyvą, kaip ji taikoma jūsų šalyje
- Jei ne, numatykite atvirą ir negriežtą licenciją
- Jei privatumas yra problema, suplanuokite reikiamas procedūras

– Priežiūra / išsaugojimas


- Dažniausiai geriausias pasirinkimas yra bendradarbiauti su duomenų centru
- Sužinokite, kaip ELRC gali jums padėti
- Taip pat galima naudoti nacionalinį atvirųjų duomenų portalą
- Duomenų laikymas žiniatinklyje nėra optimalus pasirinkimas



- Pažįstate savo duomenis
 - matomi / nematomi
- Dažnai archyvų, giliojo žiniatinklio ir t. t. negalima pasiekti iš išorės.
- Ne visiems duomenims jau taikoma PSI arba negriežta licencija
- Prieiga prie išvestinių formatų (pvz., PDF) yra mažiau efektyvi nei prieiga prie vidinių šaltinių turinio saugyklų.



Languages — the heart of
Multilingual Europe





Search


Filter by:

- ▶ Language Name
- ▶ Language Script
- ▶ Resource Type
- ▶ Availability
- ▶ Licence
- ▶ Conditions of Use
- ▶ Linguality Type
- ▶ MIME Type
- ▶ Domain

Resource Type:


Corpus: 

Lexical/Conceptual: 












Language Description: 


3 Language Resources

Order by: Resource Name A-Z ▼


 **INTERA Corpus - the Greek part of the EL-EN pair** 0 19


English Modern Greek (1463-)

 ODbL
  ODC-BY
  LO-OL-FR
  DL-DE-BY
  IODL-IT
  OGL-UK
  NLS OD-L-FIN
  NCGL-UK
  DL-DE-ZERO
  difi NL-OD-NO
  PDDL




 **my movies** 2 8

Danish Modern Greek (1463-)


 PDDL

 **press releases of the mkouts Ministry** 5 9

Croatian

 CC BY-NC
  PUBLIC DOMAIN
  PDDL

The European Language Resource Coordination - Connecting Europe Facility repository
created on the basis of the META-SHARE repository software



CC-BY-NC-SA 4.0



Helpdesk

Got a question? We're here to help!

We are happy to answer any questions on the technical or legal aspects related to the use, production, collection, processing, and sharing of language resources.

Please feel free to contact us through one of the following channels:

a Web forum

[Web forum](#)

Telephone*

+33 970 440 522

reach the Secretariat Support at

+49 681-8575 5285

Skype

CEF-AT-Helpdesk

E-mail

help@cef-at-helpdesk.org