Millaista aineistoa tarvitaan?

Mietta Lennes
Nykykielten laitos / FIN-CLARIN
Helsingin yliopisto

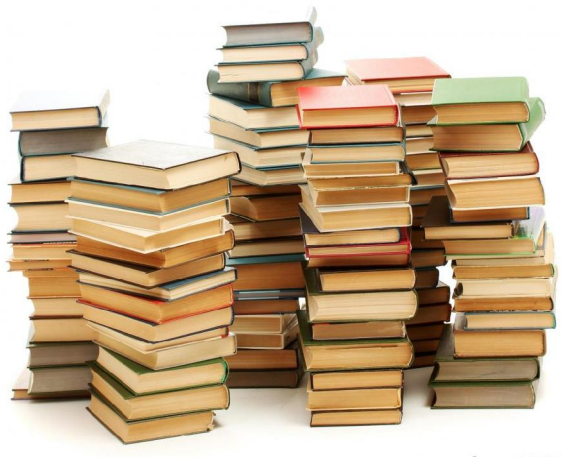
Millaista dataa? Käännöksiä

- Aineisto- eli datalähtöinen lähestymistapa
  - § Konekäännösjärjestelmät oppivat olemassa olevasta aineistosta.
  - § ELRC:n painopiste: Dataa tarvitaan kaikilla EU-kielillä
- Kieliresursseja eli ***kielivaroja*** tuotetaan dokumenteista ja muusta aineistosta.
  - § Omista tai tiedossasi olevista aineistoista voi olla suurta hyötyä!

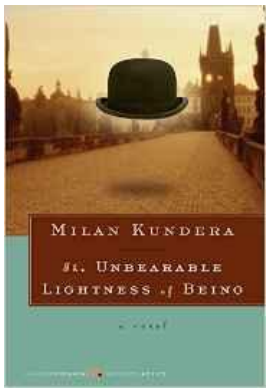
Mikä kelpaa aineistoksi?

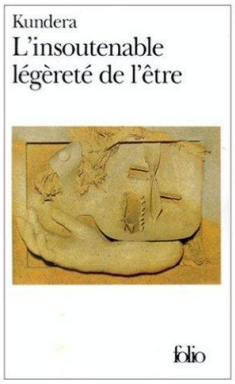
- Mitä tahansa kirjoitetussa muodossa olevaa, joka sisältää “sanoja”, vielä mieluummin “virkkeitä” ja erityisen mieluusti useilla kielillä ilmaistuja “virkkeitä”, esim.
  - raportit
  - puheet
  - verkkosivujen sisältö
  - esitteet, jne.
- “Sanojen” ja “virkkeiden” kokoelmat

Mikä kelpaa aineistoksi?



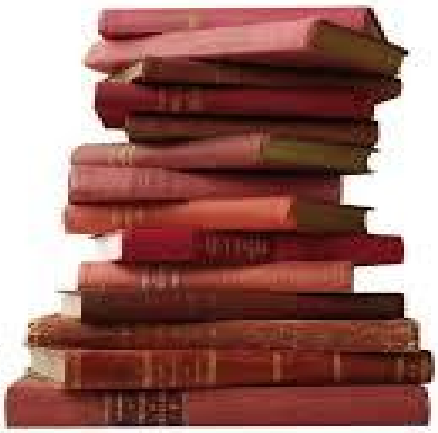
Millaista dataa? Käännöksiä





"Kohdistetut" käännökset

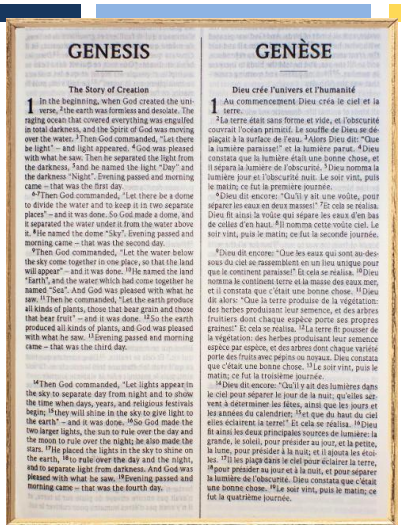
suomi



ranska

European Language Resource Coordination
Connecting Europe Facility

What types of data? “Aligned” Translation



GENESIS

The Story of Creation

1 In the beginning, when God created the universe, 2 the earth was formless and desolate. The raging ocean that covered everything was engulfed in total darkness, and the Spirit of God was moving over the water. 3 Then God commanded, "Let there be light" – and light appeared. 4 God was pleased with what he saw. Then he separated the light from the darkness, 5 and he named the light "Day" and the darkness "Night". Evening passed and morning came – that was the first day.

6-7 Then God commanded, "Let there be a dome to divide the water and to keep it in two separate places" – and it was done. So God made a dome, and it separated the water under it from the water above it. 8 He named the dome "Sky". Evening passed and morning came – that was the second day.

9 Then God commanded, "Let the water below the sky come together in one place, so that the land will appear" – and it was done. 10 He named the land "Earth", and the water which had come together he named "Sea". And God was pleased with what he saw. 11 Then he commanded, "Let the earth produce all kinds of plants, those that bear grain and those that bear fruit" – and it was done. 12 So the earth produced all kinds of plants, and God was pleased with what he saw. 13 Evening passed and morning came – that was the third day.

14 Then God commanded, "Let lights appear in the sky to separate day from night and to show the time when days, years, and religious festivals begin; 15 they will shine in the sky to give light to the earth" – and it was done. 16 So God made the two larger lights, the sun to rule over the day and the moon to rule over the night; he also made the stars. 17 He placed the lights in the sky to shine on the earth, 18 to rule over the day and the night, and to separate light from darkness. And God was pleased with what he saw. 19 Evening passed and morning came – that was the fourth day.

GENÈSE

Dieu crée l'univers et l'humanité

1 Au commencement Dieu créa le ciel et la terre.
2 La terre était sans forme et vide, et l'obscurité couvrait l'océan primitif. Le souffle de Dieu se déplaçait à la surface de l'eau. 3 Alors Dieu dit: "Que la lumière paraisse!" et la lumière parut. 4 Dieu constata que la lumière était une bonne chose, et il sépara la lumière de l'obscurité. 5 Dieu nomma la lumière jour et l'obscurité nuit. Le soir vint, puis le matin; ce fut la première journée.

6 Dieu dit encore: "Qu'il y ait une voûte, pour séparer les eaux en deux masses!" 7 Et cela se réalisa. Dieu fit ainsi la voûte qui sépare les eaux d'en bas de celles d'en haut. 8 Il nomma cette voûte ciel. Le soir vint, puis le matin; ce fut la seconde journée.

9 Dieu dit encore: "Que les eaux qui sont au-dessous du ciel se rassemblent en un lieu unique pour que le continent paraisse!" Et cela se réalisa. 10 Dieu nomma le continent terre et la masse des eaux mer; et il constata que c'était une bonne chose. 11 Dieu dit alors: "Que la terre produise de la végétation: des herbes produisant leur semence, et des arbres fruitiers dont chaque espèce porte ses propres graines!" Et cela se réalisa. 12 La terre fit pousser de la végétation: des herbes produisant leur semence espèce par espèce, et des arbres dont chaque variété porte des fruits avec pépins ou noyaux. Dieu constata que c'était une bonne chose. 13 Le soir vint, puis le matin; ce fut la troisième journée.

14 Dieu dit encore: "Qu'il y ait des lumières dans le ciel pour séparer le jour de la nuit; qu'elles servent à déterminer les fêtes, ainsi que les jours et les années du calendrier; 15 et que du haut du ciel elles éclairent la terre!" Et cela se réalisa. 16 Dieu fit ainsi les deux principales sources de lumière: la grande, le soleil, pour présider au jour, et la petite, la lune, pour présider à la nuit; et il ajouta les étoiles. 17 Il les plaça dans le ciel pour éclairer la terre, 18 pour présider au jour et à la nuit, et pour séparer la lumière de l'obscurité. Dieu constata que c'était une bonne chose. 19 Le soir vint, puis le matin; ce fut la quatrième journée.

European Language Resource Coordination
Connecting Europe Facility

Vertailukelpoiset kokoelmat

englanti

Telecommunication occurs when the exchange of information between two or more entities (communication) includes the use of technology.

Communication technology uses channels to transmit information (as electrical signals), either over a physical medium (such as signal cables), or in the form of electromagnetic waves.

The word is often used in its plural form, telecommunications, because it involves many different technologies.

suomi

Televiestintä eli kaukoviestintä tarkoittaa apuvälineiden välittämien signaalien avulla tapahtuvaa viestintää. Nykyisin televiestinnällä tarkoitetaan yleensä tietoliikennettä eli digitaalista sähköistä viestintää. Sähköisen viestinnän varhaisin muoto oli 1800-luvun alkupuolella kehitetty lennätin. Nykyisin tärkeimpiä televiestinnän muotoja ovat televisio, radio, puhelin ja internet. Televiestintä on yhä tärkeämpi osa maailmantaloutta; vuonna 2006 televiestintäteollisuuden kokonaisliikevaihto oli arviolta 1,2 triljoonaa Yhdysvaltain dollaria.

espanja

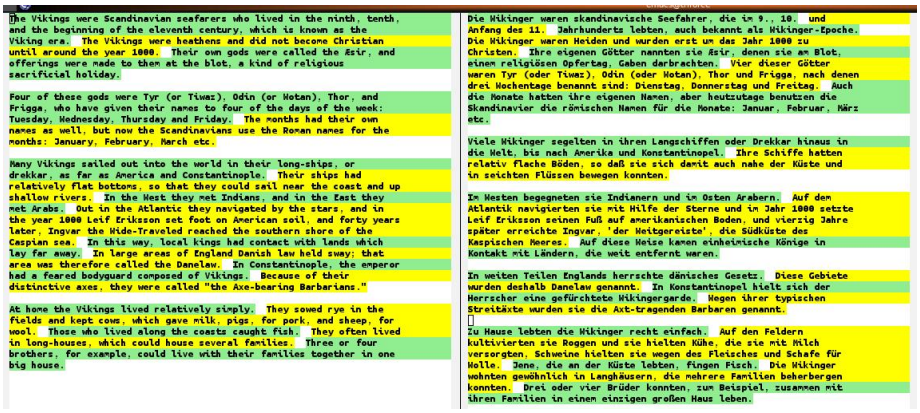
Una telecomunicación es toda transmisión y recepción de señales de cualquier naturaleza, típicamente electromagnéticas, que contengan signos, sonidos, imágenes o, en definitiva, cualquier tipo de información que se desee comunicar a cierta distancia.

Por metonimia, también se denomina telecomunicación (o telecomunicaciones, indistintamente) a la disciplina que estudia, diseña, desarrolla y explota aquellos sistemas que permiten dichas comunicaciones; de forma análoga, la ingeniería de telecomunicaciones resuelve los problemas técnicos asociados a esta disciplina.

Lähde: Ensimmäiset virkkeet televiestintää käsittelevissä englannin-, suomen- ja espanjankielisissä Wikipedia-artikkeleissa

European Language Resource Coordination
Connecting Europe Facility

Millaista dataa? “Kohdistettu” käännös



The Vikings were Scandinavian seafarers who lived in the ninth, tenth, and the beginning of the eleventh century, which is known as the Viking era. The Vikings were heathens and did not become Christian until around the year 1000. Their own gods were called the Æsir, and offerings were made to them at the blot, a kind of religious sacrificial holiday.

Four of these gods were Tyr (or Tiwaz), Odin (or Wotan), Thor, and Frigga, who have given their names to four of the days of the week: Tuesday, Wednesday, Thursday and Friday. The months had their own names as well, but now the Scandinavians use the Roman names for the months: January, February, March etc.

Many Vikings sailed out into the world in their long-ships, or drakkar, as far as America and Constantinople. Their ships had relatively flat bottoms, so that they could sail near the coast and up shallow rivers. In the West they met Indians, and in the East they met Arabs. Out in the Atlantic they navigated by the stars, and in the year 1000 Leif Eriksson set foot on American soil, and forty years later, Ingvar the Wide-Traveled reached the southern shore of the Caspian sea. In this way, local kings had contact with lands which lay far away. In large areas of England Danish law held sway; that area was therefore called the Danelaw. In Constantinople, the emperor had a feared bodyguard composed of Vikings. Because of their distinctive axes, they were called "the Axe-bearing Barbarians."

At home the Vikings lived relatively simply. They sowed rye in the fields and kept cows, which gave milk, pigs, for pork, and sheep, for wool. Those who lived along the coasts caught fish. They often lived in long-houses, which could house several families. Three or four brothers, for example, could live with their families together in one big house.

Die Wikinger waren skandinavische Seefahrer, die im 9., 10. und Anfang des 11. Jahrhunderts lebten, auch bekannt als Wikinger-Epoche. Die Wikinger waren Heiden und wurden erst um das Jahr 1000 zu Christen. Ihre eigenen Götter nannten sie Æsir, denen sie am Blot, einem religiösen Opfertag, Gaben darbrachten. Vier dieser Götter waren Tyr (oder Tiwaz), Odin (oder Wotan), Thor und Frigga, nach denen drei Wochentage benannt sind: Dienstag, Donnerstag und Freitag. Auch die Monate hatten ihre eigenen Namen, aber heutzutage benutzen die Skandinavier die römischen Namen für die Monate: Januar, Februar, März etc.

Viele Wikinger segelten in ihren Langschiffen oder Drekkar hinaus in die Welt, bis nach Amerika und Konstantinopel. Ihre Schiffe hatten relativ flache Böden, so daß sie sich damit auch nahe der Küste und in seichten Flüssen bewegen konnten.

Im Westen begegneten sie Indianern und im Osten Arabern. Auf dem Atlantik navigierten sie mit Hilfe der Sterne und im Jahr 1000 setzte Leif Eriksson seinen Fuß auf amerikanischen Boden, und vierzig Jahre später erreichte Ingvar, 'der Weitgereiste', die Südküste des Kaspischen Meeres. Auf diese Weise kamen einheimische Könige in Kontakt mit Ländern, die weit entfernt waren.

In weiten Teilen Englands herrschte dänisches Gesetz. Diese Gebiete wurden deshalb Danelaw genannt. In Konstantinopel hielt sich der Herrscher eine gefürchtete Wikingergarde. Wegen ihrer typischen Streitäxte wurden sie die Axt-tragenden Barbaren genannt.

Zu Hause lebten die Wikinger recht einfach. Auf den Feldern kultivierten sie Roggen und sie hielten Kühe, die sie mit Milch versorgten, Schweine hielten sie wegen des Fleisches und Schafe für Wolle. Jene, die an der Küste lebten, fingen Fisch. Die Wikinger wohnten gewöhnlich in Langhäusern, die mehrere Familien beherbergen konnten. Drei oder vier Brüder konnten, zum Beispiel, zusammen mit ihren Familien in einen einzigen großen Haus leben.

European Language Resource Coordination
Connecting Europe Facility

Sanakirjat / terminologiat / ontologiat

previous level in time or space.

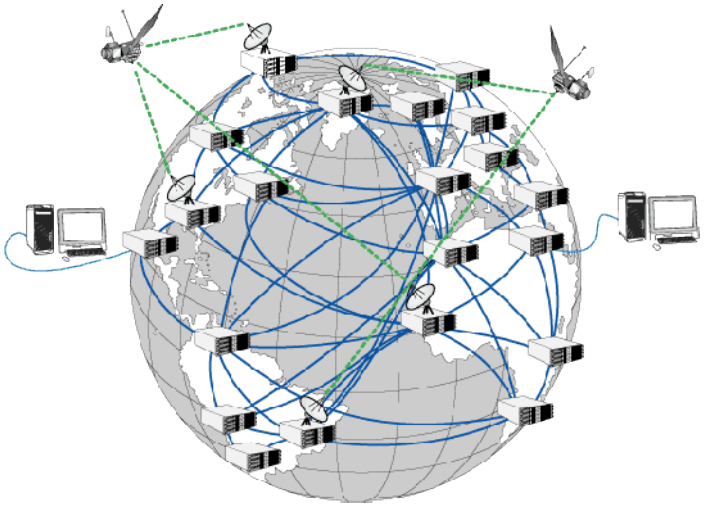
ID	FR	ES	EL
6905	abandon scolaire	abandono escolar	διακοπή της σχολικής φοίτησης
920	abats	despojo	παραπροϊόντα σφαγίων
1857	abattage d'animaux	sacrificio de animales	σφαγή ζώων
6621	abrogation	derogación	κατάργηση
5075	Abruzzes	Abruzos	Αβρουζία
5339	absentéisme	absentismo	συστηματική απουσία από την εργασία
5984	abstentionnisme	abstencionismo	αποχή
2	abus de confiance	abuso de confianza	απιστία
06	abus de droit	abuso de derecho	κατάχρηση δικαιώματος
	abus de pouvoir	abuso de poder	κατάχρηση εξουσίας
	accès à l'éducation	acceso a la educación	πρόσβαση στην εκπαίδευση
	accès à l'emploi	acceso al empleo	πρόσβαση στην αγορά εργασίας

Millaista dataa? “Kohdistettu” käännös

englanti

ranska

Mistä dataa löytyy? Digitaalinen maailma



Mikä formaatti tarvitaan? Digitaalinen tekstimuotoinen aineisto

Dokumentoitu aineisto (kuvailutiedot eli metadata)

Dublin Core Metadata Element Set

1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights

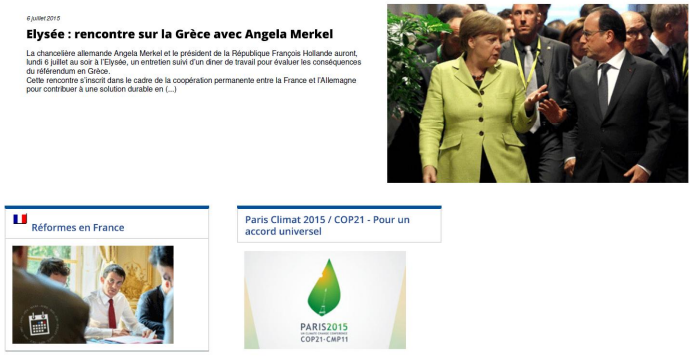
Kuinka kieliaineiston kokoaminen tapahtuu

- Raa'asta lähdemateriaalista käännösaineistoksi
  - Etsi ja tunnista lähteet
  - IPR-oikeuksien selvitys, datan keruu (esim. lataaminen tai haravointi)
  - Datan siivoaminen (poistetaan "boilerplate", "templates", kuvat, html-tagit jne., konvertoidaan)
  - Käytettävissä automaattisia työkaluja (esim. Boilerpipe)
  - Dokumentoidaan aineisto
  - Kohdistetaan käännökset aina kun tunnistettavissa; aineiston virkkeistäminen
  - Lasketaan kohdistuksen luotettavuus
  - Jaetaan aineisto

How LRs are produced

La France en Allemagne

It may start with web sites



European Language Resource Coordination
Connecting Europe Facility

Frankreich in Deutschland

6. Juli 2015

Griechenland: Staatspräsident Hollande empfängt Bundeskanzlerin (...)

Staatspräsident François Hollande empfängt am Abend des 6. Juli Bundeskanzlerin Angela Merkel im Elysée-Palast zu einem Gespräch und einem Arbeitsessen, um die Konsequenzen aus dem Referendum in Griechenland zu erörtern.
Das Treffen findet im Rahmen der ständigen Zusammenarbeit zwischen Frankreich und Deutschland mit dem Ziel statt, zu einer dauerhaften Lösung für Griechenland zu (...)

Reformagenda

Klimakonferenz Paris 2015

European Language Resource Coordination
Connecting Europe Facility

How to identify the Comparability features

Griechenland: Staatspräsident Hollande empfängt Bundeskanzlerin Merkel [fr]
Drucken
Google
Facebook
Twitter
Staatspräsident François Hollande empfängt am Abend des 6. Juli Bundeskanzlerin Angela Merkel im Elysée-Palast zu einem Gespräch und einem Arbeitsessen, um die Konsequenzen aus dem Referendum in Griechenland zu erörtern.
Das Treffen findet im Rahmen der ständigen Zusammenarbeit zwischen Frankreich und Deutschland mit dem Ziel statt, zu einer dauerhaften Lösung für Griechenland zu kommen..
Letzte Änderung 06/07/2015
Seitenanfang

- **Elysée : rencontre sur la Grèce avec Angela Merkel [de]**
- Imprimer
- Google
- Facebook
- Twitter
- La chancelière allemande Angela Merkel et le président de la République François Hollande auront, lundi 6 juillet au soir à l'Elysée, un entretien suivi d'un diner de travail pour évaluer les conséquences du référendum en Grèce.
- Cette rencontre s'inscrit dans le cadre de la coopération permanente entre la France et l'Allemagne pour contribuer à une solution durable en Grèce.
- Dernière modification : 06/07/2015
- Haut de page

Ø RAW DATA TO PROCESS

How to identify the Comparability features

Griechenland: Staatspräsident Hollande empfängt Bundeskanzlerin Merkel

François Hollande empfängt am Abend des 6. Juli Bundeskanzlerin Angela Merkel im Elysée-Palast zu einem Gespräch und einem Arbeitsessen, um die Konsequenzen aus dem Referendum in Griechenland zu erörtern.

Das Treffen findet im Rahmen der ständigen Zusammenarbeit zwischen Frankreich und Deutschland mit dem Ziel statt, zu einer dauerhaften Lösung für Griechenland zu kommen..

Letzte Änderung 06/07/2015

Seitenanfang

- **Elysée : rencontre sur la Grèce avec Angela Merkel**
- La chancelière allemande Angela Merkel et le président de la République François Hollande auront, lundi 6 juillet au soir à l'Elysée, un entretien suivi d'un diner de travail pour évaluer les conséquences du référendum en Grèce.
- Cette rencontre s'inscrit dans le cadre de la coopération permanente entre la France et l'Allemagne pour contribuer à une solution durable en Grèce.
- Dernière modification : 06/07/2015
- Haut de page

Ø RAW DATA TO PROCESS

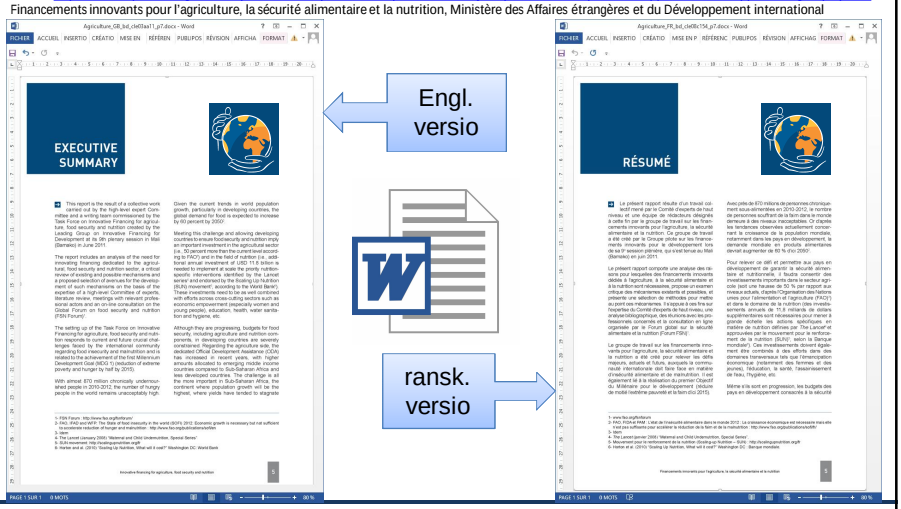
A Language Resource Factory

- How can this process be turned into a factory of LR production (Automation of the Procedure)
- Some simple illustrations
- We rather start from the Digital word
  - OCR may be considered for the less-resourced languages

European Language Resource Coordination
Connecting Europe Facility

Kaksikielisen aineiston hallinta: esimerkki (1/4)

Word docs from http://www.diplomatie.gouv.fr/fr/photos-videos-publications/publications/enjeux-planetaires-cooperation/rapports/article/rapports-du-groupe-pilote, Financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition, Ministère des Affaires étrangères et du Développement international



European Language Resource Coordination
Connecting Europe Facility

Kaksikielisen aineiston hallinta: esimerkki (2/4)

EXECUTIVE SUMMARY

This report is the result of a collective work carried out by the high-level expert Committee and a writing team commissioned by the Task Force on Innovative Financing for agriculture, food security and nutrition created by the Leading Group on Innovative Financing for Development at its 9th plenary session in Mali (Bamako) in June 2011.

The report includes an analysis of the need for innovating financing dedicated to the agricultural, food security and nutrition sector, a critical review of existing and possible mechanisms and a proposed selection of avenues for the development of such mechanisms on the basis of the

RÉSUMÉ

Le présent rapport résulte d'un travail collectif mené par le Comité d'experts de haut niveau et une équipe de rédacteurs désignés à cette fin par le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition. Ce groupe de travail a été créé par le Groupe pilote sur les financements innovants pour le développement lors de sa 9e session plénière, qui s'est tenue au Mali (Bamako) en juin 2011.

Le présent rapport comporte une analyse des raisons pour lesquelles des financements innovants dédiés à l'agriculture, à la sécurité alimentaire et à la nutrition sont nécessaires, propose un examen critique des mécanismes existants et possibles, et

Kaksikielisen aineiston hallinta: esimerkki (3/4)

Englanninkielinen versio: Raakateksti	Ranskankielinen versio: Raakateksti
Executive Summary This report is the result of a collective work carried out by the high-level expert Committee and a writing team commissioned by the Task Force on Innovative Financing for agriculture, food security and nutrition created by the Leading Group on Innovative Financing for Development at its 9th plenary session in Mali (Bamako) in June 2011. The report includes an analysis of the need for innovating financing dedicated to the agricultural, food security and nutrition sector, a critical review of existing and possible mechanisms and a proposed selection of avenues for the development of such mechanisms on the basis of the expertise of a high-level Committee of experts, literature review, meetings with relevant professional actors and an on-line consultation on the Global Forum on food security and nutrition (FSN Forum)1. The setting up of the Task Force on Innovative Financing for agriculture, food security and nutrition responds to current and future crucial challenges faced by the international community [...]	Résumé Le présent rapport résulte d'un travail collectif mené par le Comité d'experts de haut niveau et une équipe de rédacteurs désignés à cette fin par le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition. Ce groupe de travail a été créé par le Groupe pilote sur les financements innovants pour le développement lors de sa 9e session plénière, qui s'est tenue au Mali (Bamako) en juin 2011. Le présent rapport comporte une analyse des raisons pour lesquelles des financements innovants dédiés à l'agriculture, à la sécurité alimentaire et à la nutrition sont nécessaires, propose un examen critique des mécanismes existants et possibles, et présente une sélection de méthodes pour mettre au point ces mécanismes. Il s'appuie à ces fins sur l'expertise du Comité d'experts de haut niveau, une analyse bibliographique, des réunions avec les professionnels concernés et la consultation en ligne organisée par le Forum global sur la sécurité alimentaire et la nutrition (Forum FSN)1. Le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition a été créé pour relever les défis majeurs, actuels et futurs, auxquels la communauté [...]

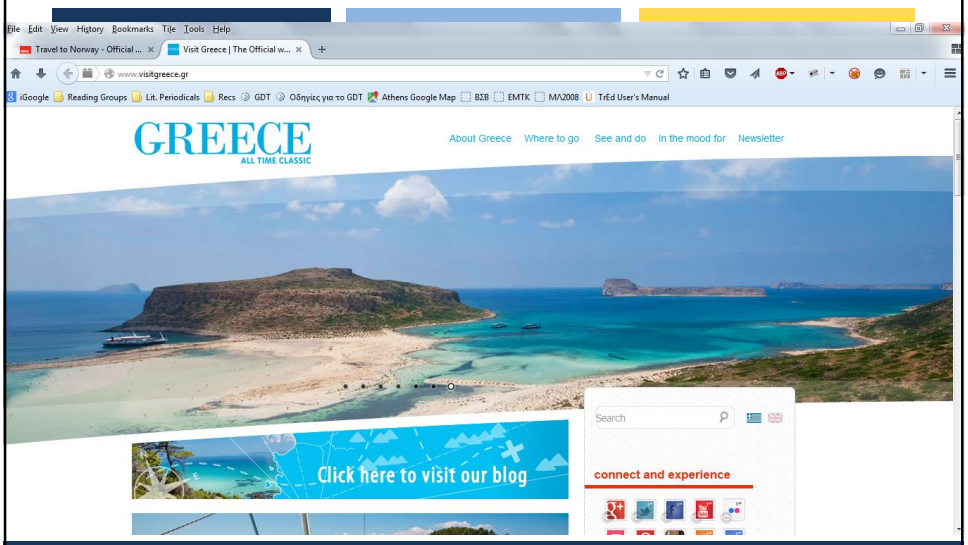
Kaksikielisen aineiston hallinta: esimerkki (4/4)

Englannin- ja ranskankielisten versioiden kohdistus:

S1. Executive Summary	**S1.** Résumé
S2. This report is the result of a collective work carried out by the high-level expert Committee and a writing team commissioned by the Task Force on Innovative Financing for agriculture, food security and nutrition created by the Leading Group on Innovative Financing for Development at its 9th plenary session in Mali (Bamako) in June 2011.	**S2.** Le présent rapport résulte d'un travail collectif mené par le Comité d'experts de haut niveau et une équipe de rédacteurs désignés à cette fin par le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition. **S3.** Ce groupe de travail a été créé par le Groupe pilote sur les financements innovants pour le développement lors de sa 9e session plénière, qui s'est tenue au Mali (Bamako) en juin 2011.
S3. The report includes an analysis of the need for innovating financing dedicated to the agricultural, food security and nutrition sector, a critical review of existing and possible mechanisms and a proposed selection of avenues for the development of such mechanisms on the basis of the expertise of a high-level Committee of experts, literature review, meetings with relevant professional actors and an on-line consultation on the Global Forum on food security and nutrition (FSN Forum)1.	**S4.** Le présent rapport comporte une analyse des raisons pour lesquelles des financements innovants dédiés à l'agriculture, à la sécurité alimentaire et à la nutrition sont nécessaires, propose un examen critique des mécanismes existants et possibles, et présente une sélection de méthodes pour mettre au point ces mécanismes. **S5.** Il s'appuie à ces fins sur l'expertise du Comité d'experts de haut niveau, une analyse bibliographique, des réunions avec les professionnels concernés et la consultation en ligne organisée par le Forum global sur la sécurité alimentaire et la nutrition (Forum FSN)1.
S4. The setting up of the Task Force on Innovative Financing for agriculture, food security and nutrition responds to current and future crucial challenges faced by the international community [...]	**S6.** Le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition a été créé pour relever les défis majeurs, actuels et futurs, auxquels la communauté [...]

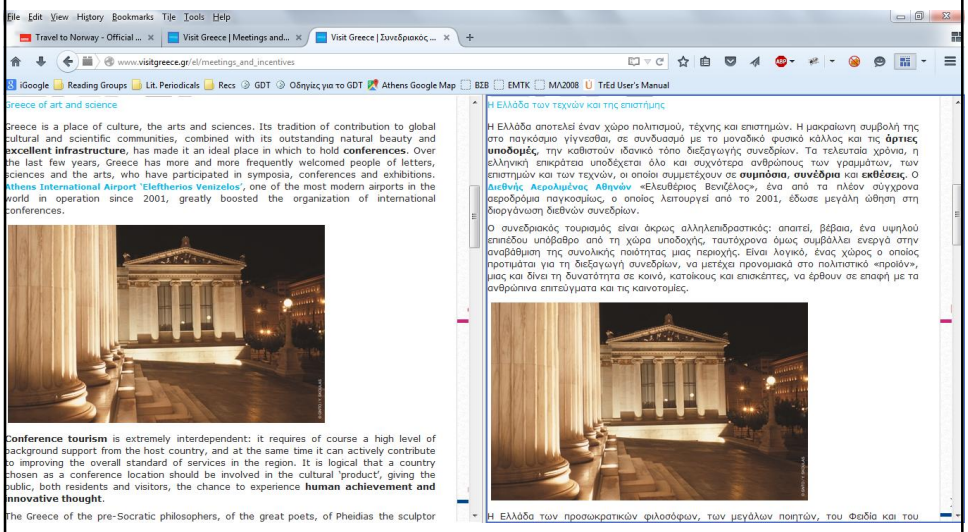
European Language Resource Coordination
Connecting Europe Facility

Many web sites are rich in multilingual content



European Language Resource Coordination
Connecting Europe Facility

How can we obtain this content…



European Language Resource Coordination
Connecting Europe Facility

Greece of art and science

Greece is a place of culture, the arts and sciences. 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**. 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. **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.

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. 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**.

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 τεχνών και της επιστήμης

Η Ελλάδα αποτελεί έναν χώρο πολιτισμού, τέχνης και επιστημών. Η μακραίωνη συμβολή της στο παγκόσμιο γίγνεσθαι, σε συνδυασμό με το μοναδικό φυσικό κάλλος και τις **άρτιες υποδομές**, την καθιστούν ιδανικό τόπο διεξαγωγής συνεδρίων. 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**. 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.

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. 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.

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.

Συνέδρια στη χώρα που γέννησε τις επιστήμες

Η Ελλάδα διαθέτει μεγάλο και υψηλής αξίας επιστημονικό δυναμικό, τόσο εντός όσο και εκτός συνόρων. Οι Έλληνες επιστήμονες, με τις εφευρέσεις, τις καινοτομίες και το ερευνητικό τους έργο πρωταγωνιστούν στη διεθνή επιστημονική κοινότητα.

Τα **επιστημονικά συνέδρια** που λαμβάνουν χώρα στην Ελλάδα είναι και πολλά και σημαντικά, αντανακλώντας τη σημασία που δίνει η χώρα στις καινοτόμες επιστήμες. Ιατρικά συνέδρια, αρχιτεκτονικά, φυσικών και ανθρωπιστικών επιστημών, πλουτίζουν την πολιτιστική ζωή της Ελλάδας, δίνοντας ταυτόχρονα τη δυνατότητα στους συνέδρους να έρθουν σε επαφή με την κοιτίδα της επιστήμης.

European Language Resource Coordination
Connecting Europe Facility

… and convert it to valuable Language Resources for Machine Translation?

File Edit View History Bookmarks Tile Tools Help

Travel to Norway - Official ... | Visit Greece | Meetings and... | Visit Greece | Συνεδριακός ... | Sentence alignment for 103.xm...

abumatran.eu/~vpapa/data/EN-EL/crawled_data/visitgreece_20150825_154605/eac25a8b-87cd-4b08-b045-571ccb003af6/xml/1234_103_u.tmx.html

iGoogle Reading Groups Lit. Periodicals Recs GDT Οδηγίες για το GDT Athens Google Map B2B EMTK ΜΛ2008 TrEd User's Manual

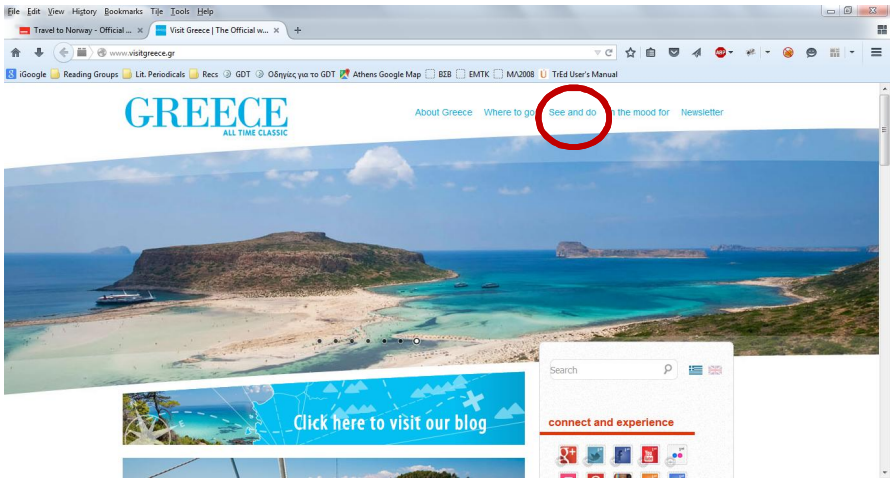
Sentence alignment for 103.xml (en) - 1234.xml (el)

#	en	el
1	Greece of art and science	Η Ελλάδα των τεχνών και της επιστήμης
2	Greece is a place of culture, the arts and sciences.	Η Ελλάδα αποτελεί έναν χώρο πολιτισμού, τέχνης και επιστημών.
3	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 μακραίωνη συμβολή της στο παγκόσμιο γίγνεσθαι, σε συνδυασμό με το μοναδικό φυσικό κάλλος και τις άρτιες υποδομές, την καθιστούν ιδανικό τόπο διεξαγωγής συνεδρίων.
4	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.
5	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.
6	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.
7	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.
8	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.
9	Scientific conferences in the land that gave birth to science	Συνέδρια στη χώρα που γέννησε τις επιστήμες
10	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.
11	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.
12	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, αντανακλώντας τη σημασία που δίνει η χώρα στις καινοτόμες επιστήμες.
13	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

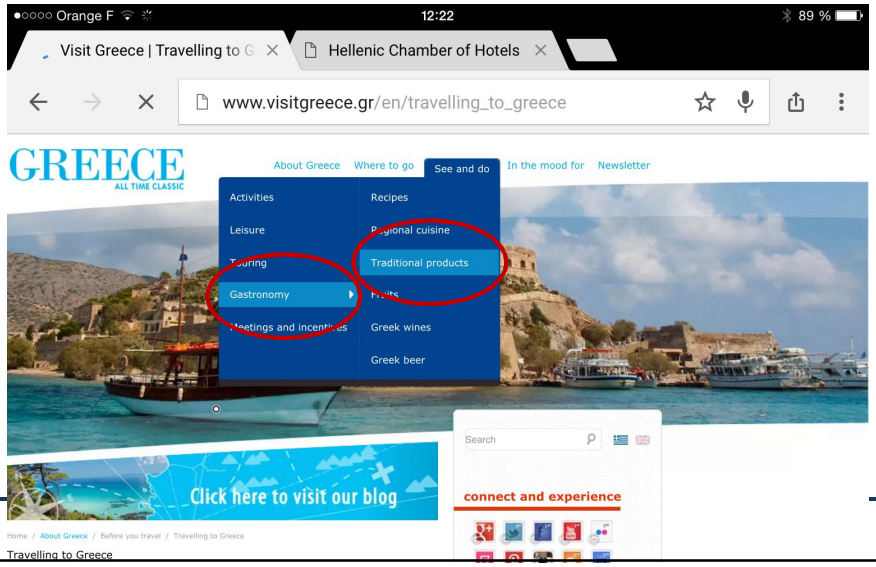
From a web page to the Factory

- How does this process scale up:
  - § Identify a “useful” source (good candidate for multilingual data)
  - § **Review and visit all the links (the URLs referenced in each page)**
    - § **“Click on each link” and move forward**
- Get each page and its “potentially” associated one in the other language
- Identify the “**domains”, “genre”,** etc. if possible
- Get rid of the “noise” (ads, format, boilerplate, etc.)
- Align (documents/files, chapters, paragraphs, sentences,)
- Check accuracy of alignment
- Use …. And share

A Journey in the meandering lines of Internet



(automatically) Follow all referenced links



31

Follow all referenced links .. Source

Line 85: ****Civilisation

Line 88: <a **href**="/en/geography">Geography

Line 91: <a **href**="/en/general_info">General Info

Line 94: <a **href**="/en/before_you_travel">Before you travel

Line 97: <a **href**="/en/greeceonthespotlight">Greece on the spotlight

Line 114: Capital

Line 121: Accommodation

Line 124: Travelling in Greece

Line 127: Travelling to Greece

Line 130: Weather

Line 133: <a href="/en/before_you_travel/passports_and_visas">Passports and Vi

referenced links … Automated process

- http://portal.elda.org/ http://portal.elda.org/en/
  - http://portal.elda.org/news/rss/
  - http://portal.elda.org/login/
    - http://portal.elda.org/en/login/
      - http://portal.elda.org/reset/
  - http://portal.elda.org/about/elra/contact/
    - http://portal.elda.org/en/about/elra/contact/
      - http://portal.elda.org/tag/85/
        - http://portal.elda.org/en/tag/85/
      - http://portal.elda.org/tag/86/
        - http://portal.elda.org/en/tag/86/

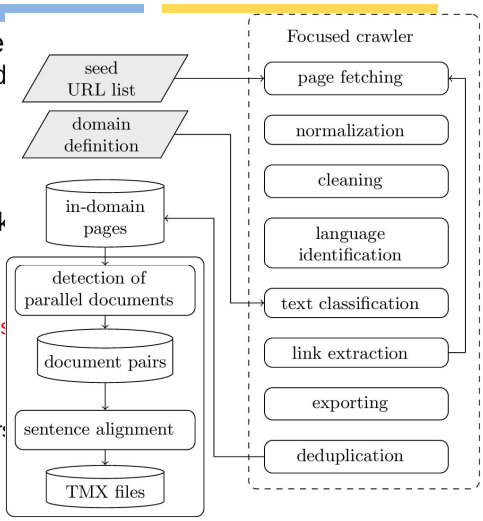
List of URLs

http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/53_50_h.tmx
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/123_144_i.tm
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/111_14_l.tmx
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/24_100_i.tmx
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/146_148_i.tm
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/1_16_i.tmx.h
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/55_51_h.tmx
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/59_48_i.tmx.
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/127_120_i.tm
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/41_103_h.tm
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/40_107_h.tm
http://nlp.ilsp.gr//ilsp-fc///mkidn_gov_20150706_175447/cec01ab3-582d-4953-80e2-90bccb862458/xml/101_46_h.tm

European Language Resource Coordination
Connecting Europe Facility

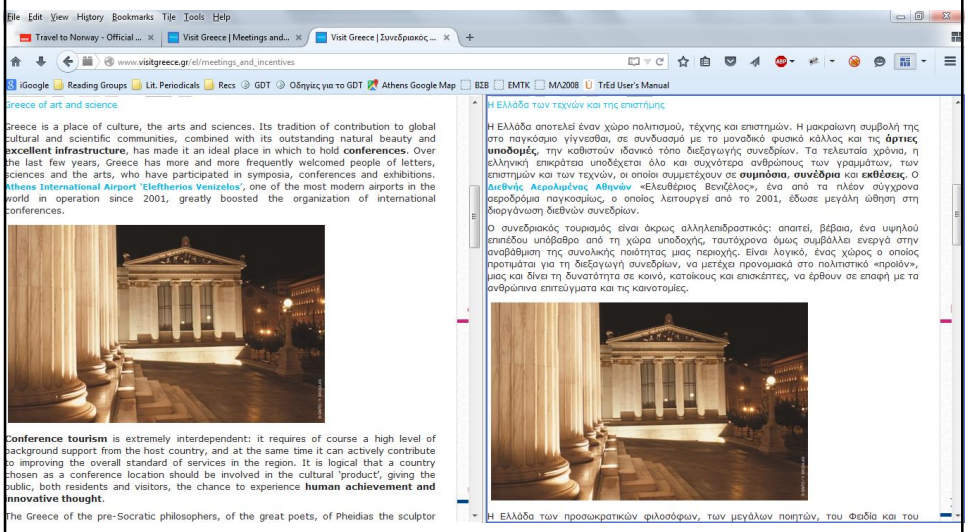
ILSP Focused Crawler

- Research prototype for acquiring ge
  or domain-specific, monolingual and
  bilingual corpora
- Input:
  - Domain definitions (lists of terms)
  - **Seed URLs**
- Modules (open source libraries/toolk
  - Page Fetching/Text Extraction
  - Normalization and Metadata Extraction
  - Boilerplate Detection (Boilerpipe)
  - Language Detection (covering > 50 langs
  - Text Classification
  - Exact and near de-duplication
  - Detection of pairs of parallel documents
  - Sentence alignment (Hunalign and others
- **Generates lists of**
  - document pairs and
  - segment pairs in TMX files



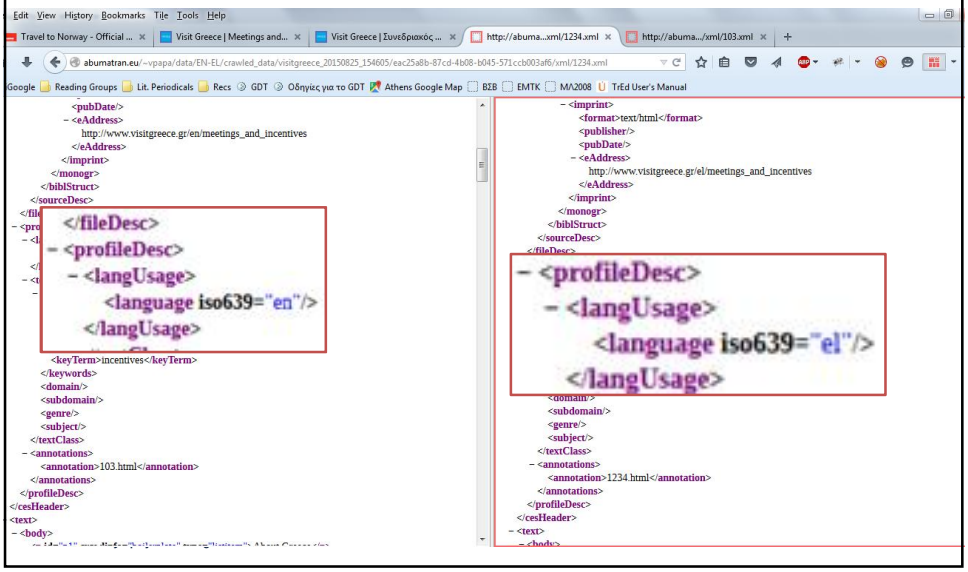
European Language Resource Coordination
Connecting Europe Facility

… it integrates technologies to crawl (part of a /multiple pages) website…



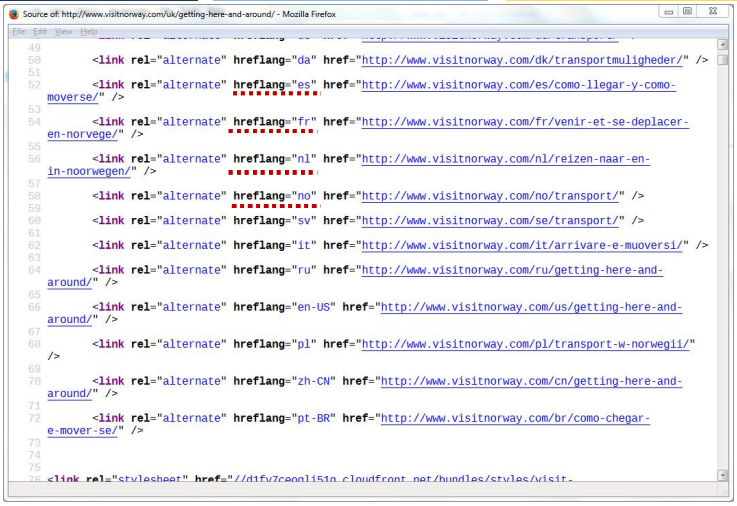
European Language Resource Coordination
Connecting Europe Facility

… identify the language of each crawled page …



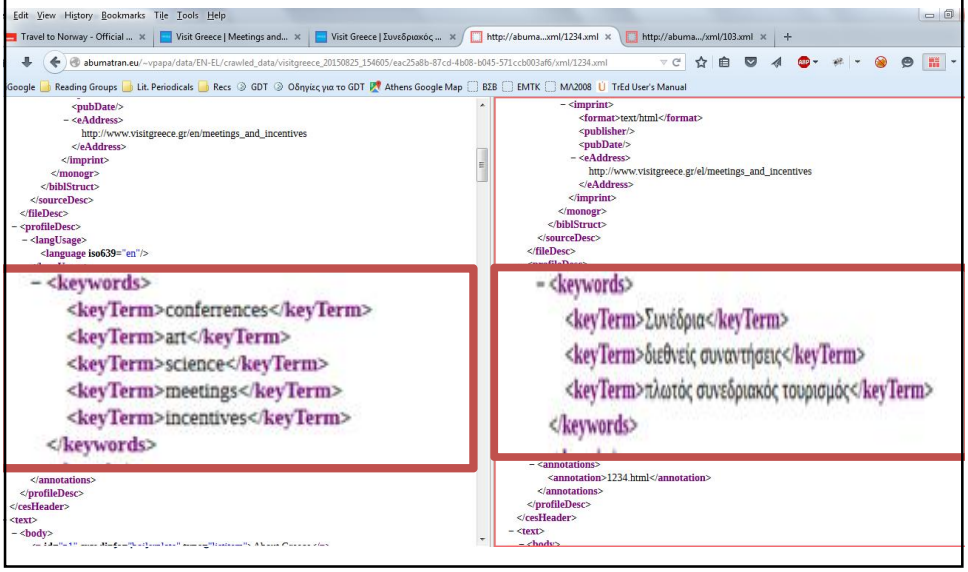
European Language Resource Coordination
Connecting Europe Facility

… identify the language of each crawled page



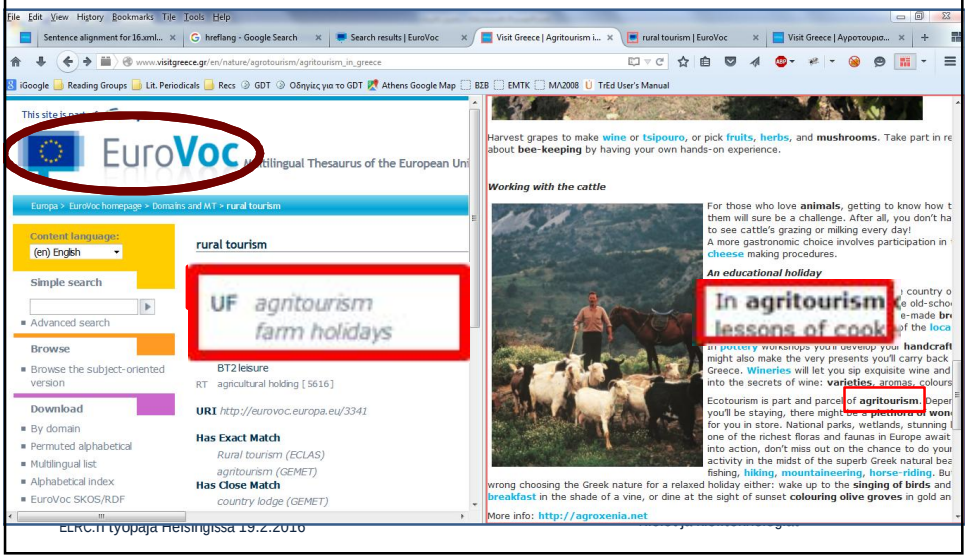
European Language Resource Coordination
Connecting Europe Facility

… extract several types of data descriptors (metadata)



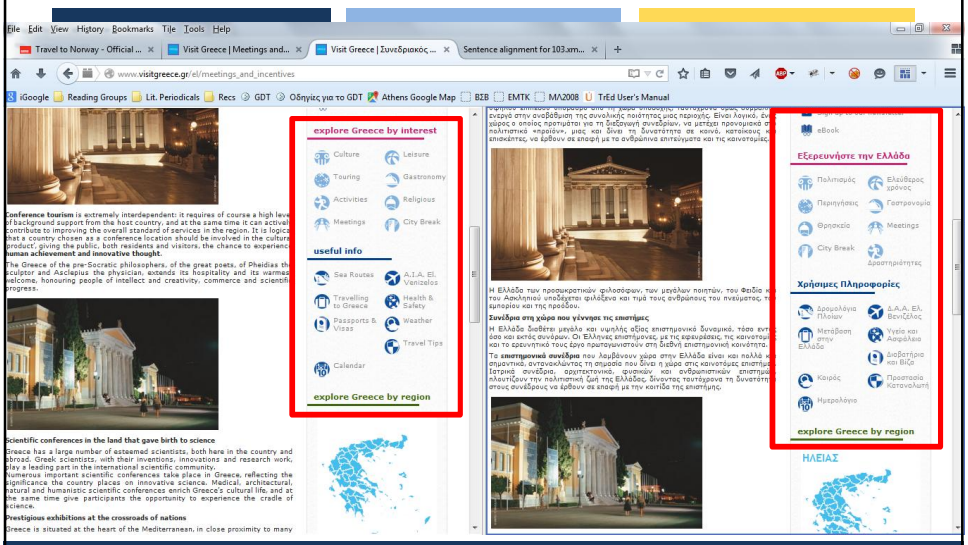
European Language Resource Coordination
Connecting Europe Facility

… and optionally classify each page as relevant or not to a user-defined domain



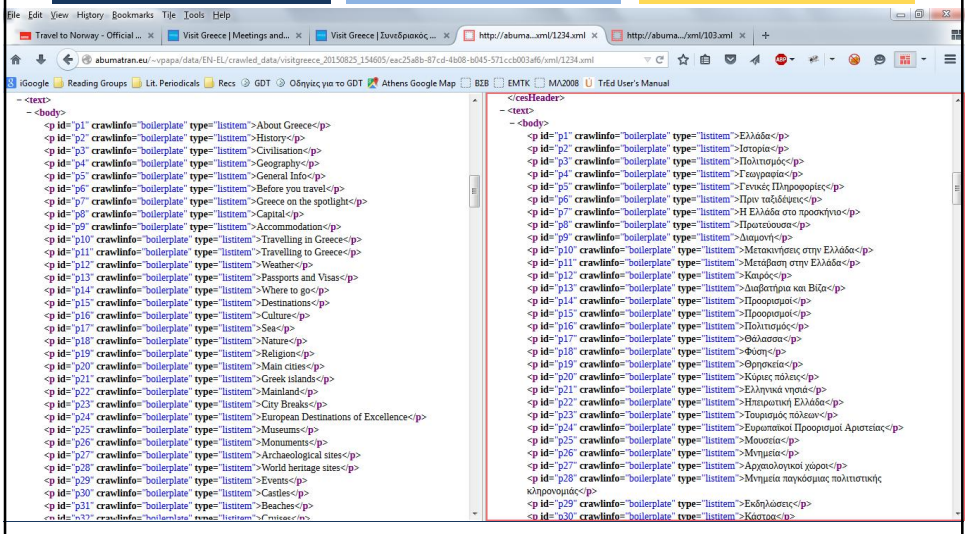
European Language Resource Coordination
Connecting Europe Facility

It can detect boilerplate text …



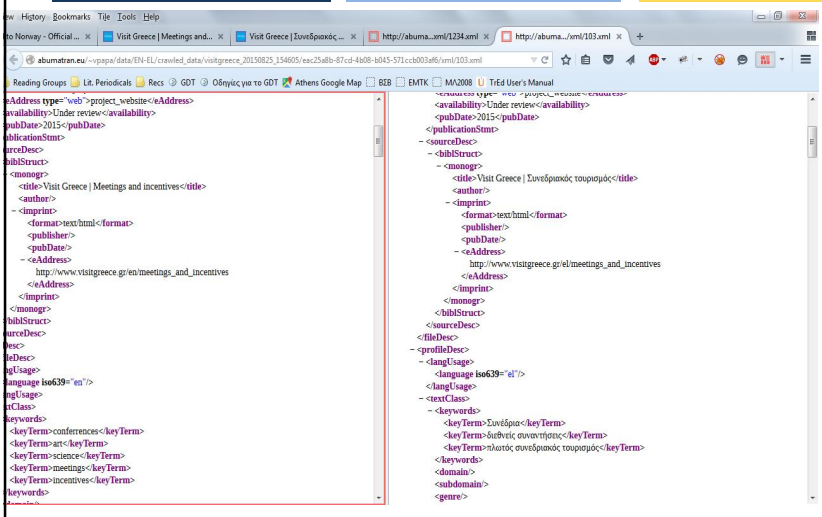
European Language Resource Coordination
Connecting Europe Facility

… and makes it easy for LR developers to filter it out



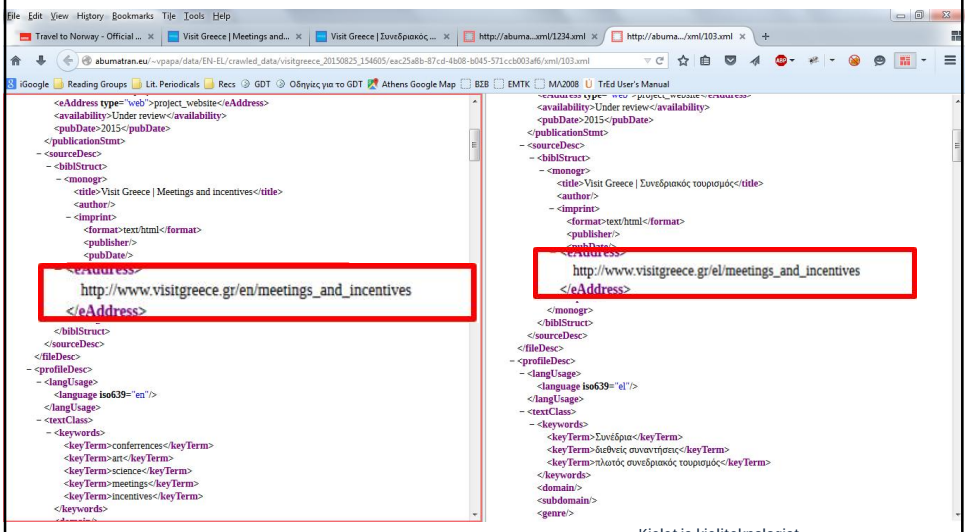
European Language Resource Coordination
Connecting Europe Facility

… HTML structure and/or URL similarity to detect document pairs



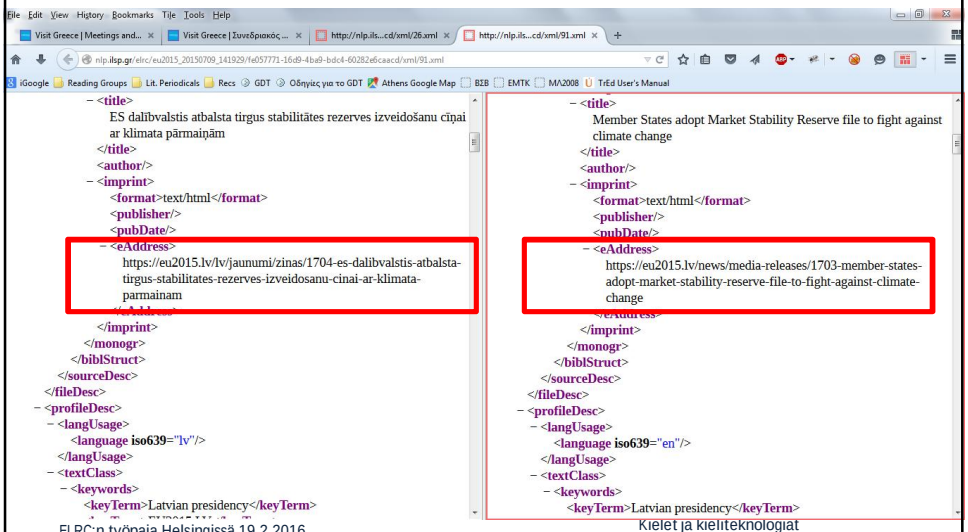
European Language Resource Coordination
Connecting Europe Facility

… HTML structure and/or URL similarity to detect document pairs



European Language Resource Coordination
Connecting Europe Facility

Sometimes URLs are not enough for finding document pairs…



European Language Resource Coordination
Connecting Europe Facility

… sentences and align it

#	en	de
1	Securing energy supplies in Europe	Energieversorgung in Europa sichern
2	Chancellor Angela Merkel and Polish Prime Minister Donald Tusk have expressed their disappointment that Russia has so far failed to implement the provisions of the Geneva accord.	Bundeskanzlerin Merkel und der polnische Ministerpräsident Tusk haben sich enttäuscht gezeigt, dass Russland die Genfer Vereinbarungen bisher nicht umgesetzt hat.
3	At their meeting in Berlin they also called for secure energy supplies in Europe.	Gleichzeitig mahnten sie bei ihrem Treffen in Berlin die Sicherung der Energieversorgung in Europa an.
4	Talks focused on the current situation in Ukraine and European energy policy Photo:	Die aktuelle Lage in der Ukraine und die europäische Energiepolitik waren Themen des Gesprächs.
5	Bundesregierung/Denzel	Foto:
6	Before their meeting at the Federal Chancellery, Chancellor Angela Merkel and Prime Minister Donald Tusk made their positions on Ukraine and Russia quite clear.	Bundesregierung/Denzel Vor ihrem Gespräch im Kanzleramt machten Bundeskanzlerin Angela Merkel und Ministerpräsident Donald Tusk ihre Position bezüglich der Ukraine und Russland deutlich.
7	Telephone call with Vladimir Putin	Telefonat mit Putin
8	The Chancellor reported that she had spoken by telephone that morning with Russian President Vladimir Putin.	Die Kanzlerin teilte mit, dass sie am Morgen mit dem russischen Präsidenten Wladimir Putin telefoniert habe.
9	She had given him to understand that she was disappointed by Russia's failure to take any	Dabei habe sie ihm zu verstehen gegeben, dass sie von der mangelhaften Umsetzung der Genfer

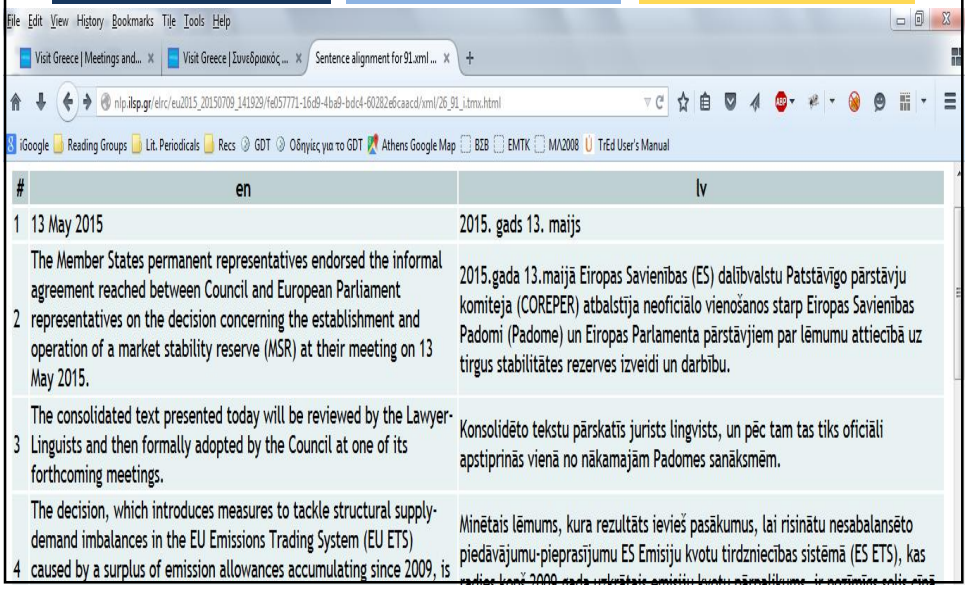
European Language Resource Coordination
Connecting Europe Facility

… sentences and align it

#	en	de
1	Reliable partnership based on trust	Vertrauensvolle und verlässliche Partnerschaft
2	"The stage has been set for further intensive cooperation," declared Chancellor Angela Merkel after her meeting with Colombian President Juan Manuel Santos.	"Die Weichen für eine weitere intensive Zusammenarbeit sind gestellt." Das hat Bundeskanzlerin Merkel nach dem Treffen mit dem kolumbianischen Präsidenten Santos erklärt.
3	The Chancellor intends to support the peace process in Colombia, partly through cooperation in the fields of research, education and climate change mitigation.	Den Friedensprozess in Kolumbien will Merkel unter anderem durch weitere Kooperationen in den Bereichen Forschung, Bildung und Klimaschutz unterstützen.
4	Colombia's President seeks support in the peace process with FARC rebels Photo:	Kolumbiens Präsident wirbt für Unterstützung beim Friedensprozess mit den FARC-Rebellen. Foto:
5	Bundesregierung/Denzel	Bundesregierung/Denzel
6	"We have a cordial and reliable partnership based on trust," declared Chancellor Angela Merkel after her meeting with Colombian President Juan Manuel Santos.	"Wir sind in einer vertrauensvollen, freundschaftlichen und verlässlichen Partnerschaft", erklärte Bundeskanzlerin Angela Merkel nach dem Treffen mit dem kolumbianischen Präsidenten Juan Manuel Santos.
7	Their talks focused on Colombia's peace process, bilateral relations and economic and regional issues.	Im Mittelpunkt des Gesprächs standen der kolumbianische Friedensprozess, die bilateralen Beziehungen sowie wirtschaftliche und regionalpolitische Themen.
8	Fostering the peace process	Friedensprozess fördern
9	"The situation today in Colombia is marked by the courageous peace process initiated by the President, which is currently in a crucial phase," said the Chancellor following talks with Colombia's President Juan Manuel Santos.	"Die aktuelle Situation in Kolumbien ist dadurch gekennzeichnet, dass der Präsident einen mutigen Friedensprozess initiiert hat, der im Augenblick in einer entscheidenden Phase ist", erläuterte die Kanzlerin nach dem Treffen mit dem kolumbischen Präsidenten Juan Manuel Santos.
10	She reported that she had pledged Germany's full support in this process.	Hierfür habe sie die volle deutsche Unterstützung zugesagt, sagte Merkel.
11	In Columbia a conflict has been smouldering for decades between right-wing paramilitary groups, left-wing guerrillas and the Colombian army.	In Kolumbien schwelt seit Jahrzehnten ein Konflikt zwischen rechtsgerichteten Paramilitärs, linksgerichteten Guerillatruppen und der kolumbianischen Armee.

European Language Resource Coordination
Connecting Europe Facility

… sentences and align it

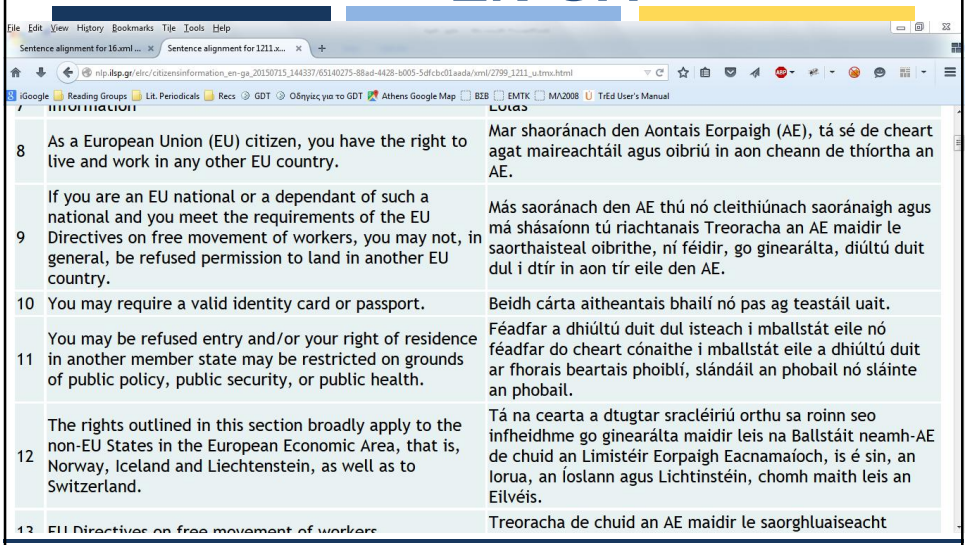


#	en	lv
1	13 May 2015	2015. gads 13. maijs
2	The Member States permanent representatives endorsed the informal agreement reached between Council and European Parliament representatives on the decision concerning the establishment and operation of a market stability reserve (MSR) at their meeting on 13 May 2015.	2015.gada 13.maijā Eiropas Savienības (ES) dalībvalstu Patstāvīgo pārstāvju komiteja (COREPER) atbalstīja neoficiālo vienošanos starp Eiropas Savienības Padomi (Padome) un Eiropas Parlamenta pārstāvjiem par lēmumu attiecībā uz tirgus stabilitātes rezerves izveidi un darbību.
3	The consolidated text presented today will be reviewed by the Lawyer-Linguists and then formally adopted by the Council at one of its forthcoming meetings.	Konsolidēto tekstu pārskatīs jurists lingvists, un pēc tam tas tiks oficiāli apstiprinās vienā no nākamajām Padomes sanāksmēm.
4	The decision, which introduces measures to tackle structural supply-demand imbalances in the EU Emissions Trading System (EU ETS) caused by a surplus of emission allowances accumulating since 2009, is	Minētais lēmums, kura rezultāts ievieš pasākumus, lai risinātu nesabalansēto piedāvājumu-pieprasījumu ES Emisiju kvotu tirdzniecības sistēmā (ES ETS), kas radies kopš 2009. gada uzkrātais emisiju kvotu pārpalikums, ir paredzēts veikt tīrā

European Language Resource Coordination
Connecting Europe Facility

Our tools supports all EU languages!

EN-GA

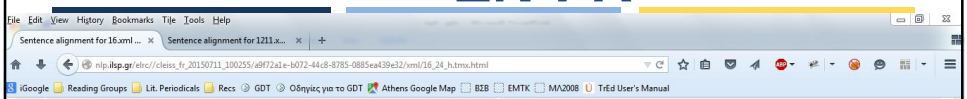


8	As a European Union (EU) citizen, you have the right to live and work in any other EU country.	Mar shaoránach den Aontais Eorpaigh (AE), tá sé de cheart agat maireachtáil agus oibriú in aon cheann de thíortha an AE.
9	If you are an EU national or a dependant of such a national and you meet the requirements of the EU Directives on free movement of workers, you may not, in general, be refused permission to land in another EU country.	Más saoránach den AE thú nó cleithiúnach saoránaigh agus má shásaíonn tú riachtanais Treoracha an AE maidir le saorthaisteal oibrithe, ní féidir, go ginearálta, diúltú duit dul i dtír in aon tír eile den AE.
10	You may require a valid identity card or passport.	Beidh cárta aitheantais bhailí nó pas ag teastáil uait.
11	You may be refused entry and/or your right of residence in another member state may be restricted on grounds of public policy, public security, or public health.	Féadfar a dhiúltú duit dul isteach i mballstát eile nó féadfar do cheart cónaithe i mballstát eile a dhiúltú duit ar fhorais beartais phoiblí, slándáil an phobail nó sláinte an phobail.
12	The rights outlined in this section broadly apply to the non-EU States in the European Economic Area, that is, Norway, Iceland and Liechtenstein, as well as to Switzerland.	Tá na cearta a dtugtar sracléiriú orthu sa roinn seo infheidhme go ginearálta maidir leis na Ballstáit neamh-AE de chuid an Limistéir Eorpaigh Eacnamaíoch, is é sin, an Iorua, an Íoslann agus Lichtinstéin, chomh maith leis an Eilvéis.
13	EU Directives on free movement of workers	Treoracha de chuid an AE maidir le saorghluaiseacht

European Language Resource Coordination
Connecting Europe Facility

it supports all EU languages!

EN-FR



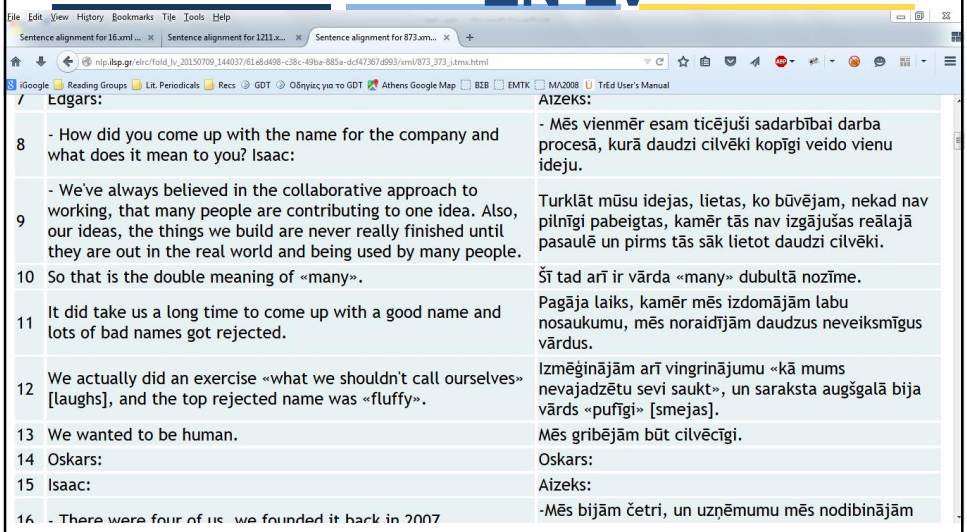
3	Sickness, maternity and paternity insurance benefits are provided in Metropolitan France by the local Health Insurance Funds (Caisses Primaires d'Assurance Maladie/ CPAM) and in the Overseas Departments by the General Social Security Funds (CGSS).	Les prestations de l'assurance maladie, maternité et paternité sont attribuées par les caisses primaires d'assurance maladie (CPAM) en métropole et par les caisses générales de sécurité sociale (CGSS) dans les départements d'outre-mer.
4	To qualify for benefits, the claimant must have paid a certain amount in contributions or worked a certain number of hours within a given reference period.	Le droit à ces prestations est subordonné soit au versement d'un certain montant de cotisations, soit à un nombre d'heures de travail durant chaque période de référence.
5	To qualify for two years' health or maternity care, the claimant must:	Pour avoir droit au remboursement des soins pendant ux ans, en cas de maladie ou de maternité, l'assuré it justifier :
6	have worked for at least 60 hours, or have paid contributions on an amount equal to at least 60 times the hourly SMIC over a period of one month;	avoir travaillé au moins 60 heures, ou avoir cotisé sur un salaire au moins égal à 60 fois le montant du SMIC horaire, pendant un mois ;
7	or have worked for at least 120 hours, or have paid contributions on an amount equal to at least 120 times the hourly SMIC over a period of three months;	ou avoir travaillé au moins 120 heures, ou avoir cotisé sur un salaire au moins égal à 120 fois le montant du SMIC horaire, pendant trois mois ;
8	or have worked at least 400 hours, or have paid contributions on an amount equal to at least 400 times the	ou avoir travaillé au moins 400 heures, ou avoir cotisé sur un salaire au moins égal à 400 fois le montant du SMIC

Score: 5.038181

European Language Resource Coordination
Connecting Europe Facility

It supports all EU languages!

EN-LV



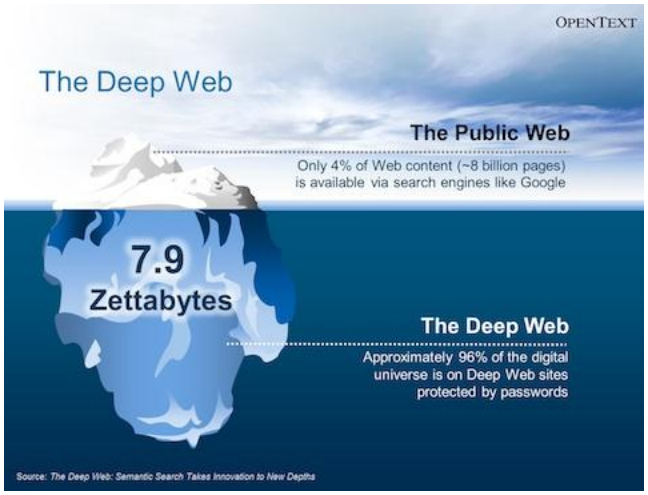
7	Edgars:	Aizeks:
8	- How did you come up with the name for the company and what does it mean to you? Isaac:	- Mēs vienmēr esam ticējuši sadarbībai darba procesā, kurā daudzi cilvēki kopīgi veido vienu ideju.
9	- We've always believed in the collaborative approach to working, that many people are contributing to one idea. Also, our ideas, the things we build are never really finished until they are out in the real world and being used by many people.	Turklāt mūsu idejas, lietas, ko būvējam, nekad nav pilnīgi pabeigtas, kamēr tās nav izgājušas reālajā pasaulē un pirms tās sāk lietot daudzi cilvēki.
10	So that is the double meaning of «many».	Šī tad arī ir vārda «many» dubultā nozīme.
11	It did take us a long time to come up with a good name and lots of bad names got rejected.	Pagāja laiks, kamēr mēs izdomājām labu nosaukumu, mēs noraidījām daudzus neveiksmīgus vārdus.
12	We actually did an exercise «what we shouldn't call ourselves» [laughs], and the top rejected name was «fluffy».	Izmēģinājām arī vingrinājumu «kā mums nevajadzētu sevi saukt», un saraksta augšgalā bija vārds «pufīgi» [smejas].
13	We wanted to be human.	Mēs gribējām būt cilvēcīgi.
14	Oskars:	Oskars:
15	Isaac:	Aizeks:
16	- There were four of us, we founded it back in 2007	-Mēs bijām četri, un uzņēmumu mēs nodibinājām

European Language Resource Coordination
Connecting Europe Facility

Parempi vaihtoehto

- Verkossa vain “näkyvä osa”: organisaatioiden sisällä paljon lisää
- Auta meitä tunnistamaan sopivia aineistolähteitä!
- Teidän tuellanne on mahdollista rakentaa käännösaineistojen tuotantolinja tai “tehdas”: siis automatisoida tekstidokumenttien keruu ja prosessointi.

Näkyvissä vs. olemassa oleva data

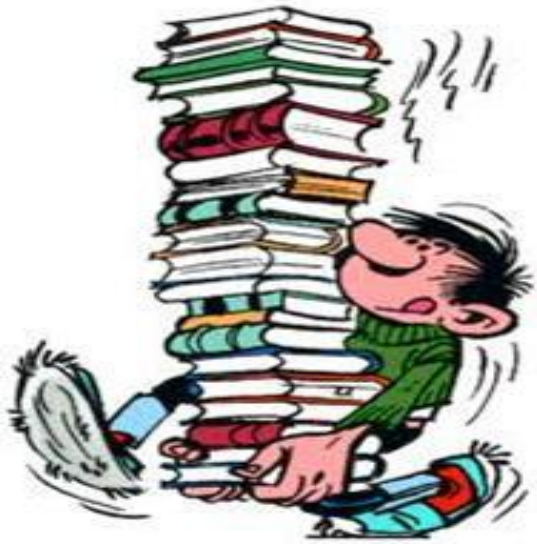


Meidän panoksemme… “Deep web”

Parempi vaihtoehto

- Dokumentteja on jo olemassa
  - erilaisissa dokumentaatiokeskuksissa (käännetyt raportit, lehtiset, esitteet, puheet, verkkosivut jne.)
  - kielipalvelujen tarjoajilla, joilta käännöspalvelut hankitaan
- Auta meitä tunnistamaan ja ottamaan yhteyttä molempiin tahoihin!

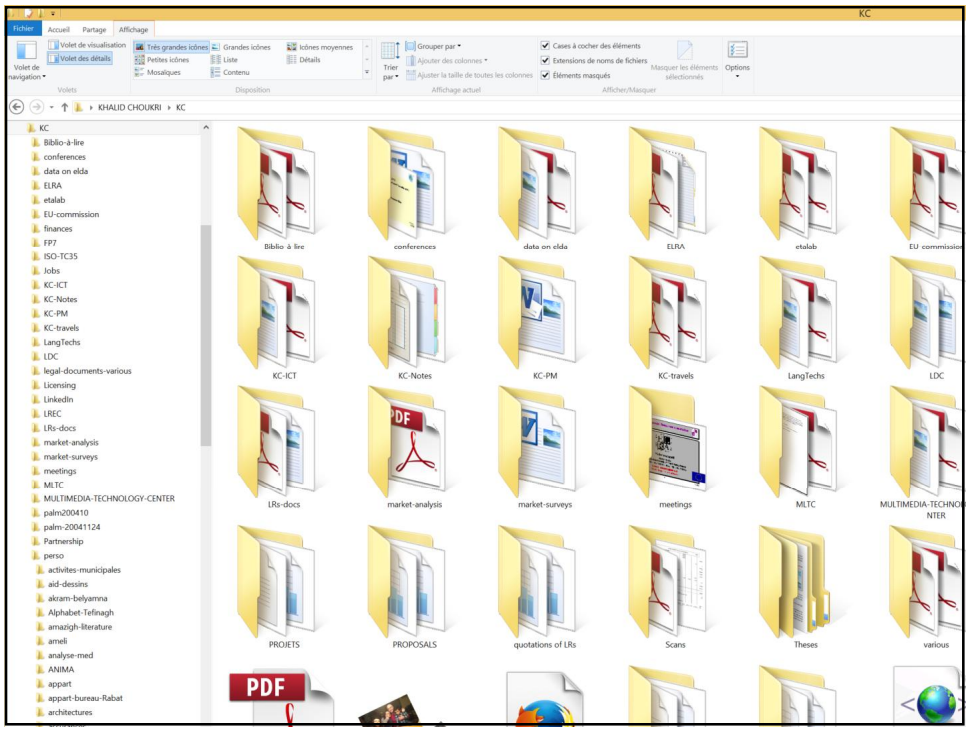
Panoksesi on välttämätön – tehdään yhteistyötä!



TUO OMAT KIELIAINEISTOSI

European Language Resource Coordination
Connecting Europe Facility

Panoksesi on välttämätön – tehdään yhteistyötä!



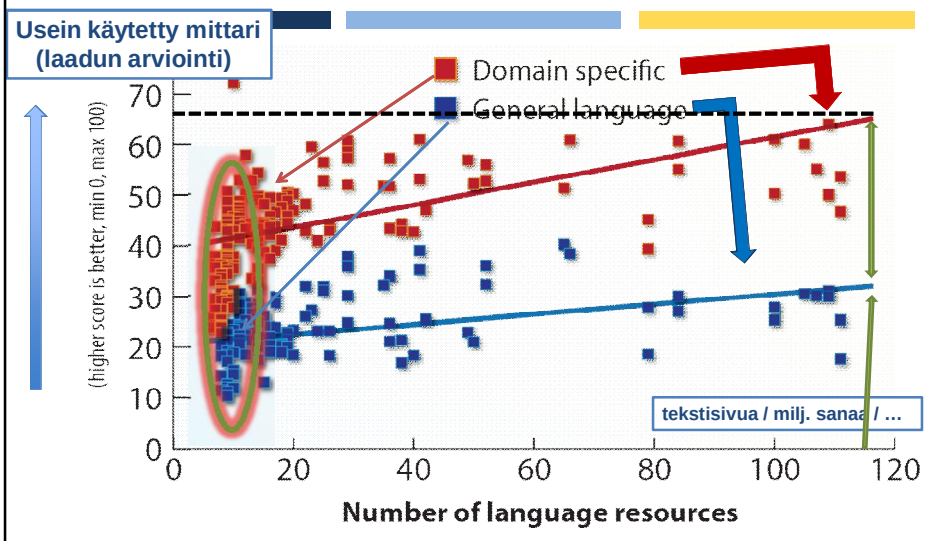
European Language Resource Coordination
Connecting Europe Facility

- Aineiston siirtämisestä ja repositorion perustamisesta tarjotaan erikseen lisäohjeita.

- **Paljonko aineistoa tarvitaan?**

European Language Resource Coordination
Connecting Europe Facility

Impact of number of language resources on Statistical MT quality



JOHTOPÄÄTÖKSET

- Aineistoa tuotetaan **paketoimalla olemassa olevia aineistoja uudelleen uutta tarkoitusta varten.**
- Tärkeää, koska datalähtöinen lähestymistapa on konekääntämisessä hyvin tehokas.

- *Aineistojemme arvoa ei pidä aliarvioida!*

- *Seuraavissa sessioissa: Kuinka voit osallistua CEF:AT:n kehittämiseen ja hyötyä siitä*