The GEROM Project Search Engine

Marco Stefano Tomatis Università degli Studi Internazionali di Roma - UNINT

Abstract

This paper aims at describing the technical engine underlying the project "GEROM", an online bilingual (German-Italian) terminological tool which has been developed by the Johannes Gutenberg - Universität of Mainz/Germersheim, Germany and Unint - University of International Studies of Rome, Italy. GEROM has been designed to facilitate the comprehension, analysis and translation of Italian and German texts dealing with different social and cultural aspects of the two Countries which are taken into account. Therefore, in order to provide users with a terminological support which focuses on examples of real language, a corpus-based approach has been adopted. Furthermore, since the core of the system is founded on a web-oriented service, a robust search interface based on CGI (Common Gateway Interface) technology has been implemented.

Key Words: terminology; CGI; search engine

Intercultural communication has always represented a complex, multifaceted issue. For this reason, it has been dealt in different ways and with different approaches by linguistic science in its broadest sense. Whereas from a theoretical point of view the gap between signifier and meaning, which represents the core of the problem, has been managed by various disciplines ranging from logics and semantics to philosophy of language, its practical evidence could be effectively faced only after the computer science provided linguists with a robust technological framework. The continuous improving of processing speed, storage space and programming languages which make pattern matching of linguistic structures an easy task provided new impulse in linguistic research, giving birth to totally new disciplines like Computational Linguistics, Corpus Linguistics, and Machine Translation. Amongst the above mentioned research fields, Corpus Linguistics is, in absolute terms, the most exploited as it is the only one that can provide scholars and researchers with data mirroring the real usage of a language. The Gerom project is developed starting from this corpus-driven approach. It aims at improve the translation of German and Italian texts by showing the real usage of language-specific terminology in a variety of contexts.

1. Linguistic data management

In order to reduce the system structure as much as possible, all linguistic data have been stored in a web server in plain text, UTF-8 format and no server-side database or data management software has been adopted. Such a choice, which may sound strange to most Information Technology professionals, has been made on purpose. Indeed, despite their renowned efficiency, in general databases require specific server technology to achieve good response performances when many concurrent search queries are entered.

Furthermore, databases require periodical maintenance sessions and serious problems may occur if this task is not regularly accomplished. Another important issue is tied to data management aspects. For security reasons and retrieval effectiveness, all databases convert plain text into an indexed, binary-coded structure which cannot be easily accessed in real time by other software products. Data exchange is possible only after a specific table export process is run on demand.

As a consequence of this, the human operator is, in fact, impeded to visualize and modify the linguistic information with a simple text editor. On the contrary, all the tables which are stored in the server can be accessed and manipulated via SQL language by the same database management system which has been used to create them. If, on the one hand, this feature may greatly improve the search engine efficiency, on the other hand it seriously limits the data structure flexibility, making changes hard to implement in case sudden and relevant modifications of the linguistic knowledge-base are required (i.e. changing the primary index or adding a field or a record in the middle of a table in use).

1.1 Information structure

Although a complete and thorough description of the linguistic dataset which has been adopted goes beyond the scope of this paper, some words will be spent on this subject. Firstly, it is important to keep in mind that the linguistic resources the system may refer to, cover two functions and differentiate accordingly. As for the role, they play the primary and most important resource deals with providing users with all the relevant terminological information tied to any German or Italian search query which is entered in the GEROM home page. To accomplish this service, the content of a particular file which works like an enriched bilingual terminological dictionary is read.

Structurally organized like a spreadsheet, this file is a tab-separated value UTF-8 plain text in which each record contains 75 different fields ranging from the terminological expression and its synonyms, to the name of the corpora where the said expression can be found in. Because of its leading function, the bilingual terminological resource is the first file the CGI engine must access. Only after the user's request has exactly matched the content of one specific field of the said terminological file, all the query-related linguistic information is extracted and properly mark upped to be displayed in HTML format. The second kind of linguistic resources, instead, is represented by all the different corpora available in the server. Each corpus is an XML-structured collection of plain text, single theme, monolingual articles which are enriched with relevant metadata information. As a final note, it is useful to remark that all the linguistic resources the system may use have been edited by Professor Claudio Fantinuoli of the Johannes Gutenberg University and Professor Fabio Proia of Unint.

2. CGI scripts

As mentioned before, the interaction between the user requests and the web server is managed by CGI technology. Without facing specific technical details, the most important feature characterizing this approach is the possibility to establish a dynamic client-server communication by adopting any suitable server-side programming language. Since the system is required to manipulate plain text strings only, the most convenient way to do this is choosing a text-oriented programming language like AWK, Perl, or the most recent Python. Despite their differences in terms of specific scope and purpose they were designed for, all of the abovementioned computer languages descend from "C" and offer a powerful backward-compatible, regexp-centered, high level, straightforward syntax. These features are of primary importance because they allow developers to easily maintain or upgrade their program by adding new functions at any time or, if needed, translate the source code from a scripting language to another with minimal adaptations.

In order to achieve a better visualization of the terminological information and improve the interactive exploitation of the online platform, two different CGI algorithms have been created. The most complex of them has been designed to manage multiple functions, so it represents the core of the system. The other one, instead, plays a secondary role. It activates on user demand only and limits to extract and display a wider area of the text the main program returns to the general response page.

2.1 Input query

For clarity reasons, the CGI script which is run for first will be described by steps, focusing on its functional workflow. As mentioned before, this program performs a number of different tasks which range from dealing with all the incoming requests from remote users, to the creation of all the response pages. Therefore, this script is automatically run by the Apache web server whenever a query is entered in the GEROM home page search box. Because the first task of the script is designed to handle the user queries, particular features of the HTTP protocol should be properly managed in order to obtain an orthographically standard input string.

Although apparently strange, this operation is mandatory because, according to the transfer protocol, any query can be transmitted to the web server in plain ASCII code only. This implies that all the accented letters and the characters with diacritical marks are automatically transformed in a non-intelligible string before being sent to the server (e.g. "à" > "%C3%A0"). Starting from this assumption, it becomes clear that in order to avoid mismatch issues, the whole user query must undergo a reconversion process before being stored in a variable for any further use. As regards queries, it is important to put in evidence that the system has been designed to accept both single and multiword terms as an input string. Moreover, to help users avoid any possible query misspelling error, a precompiled, close list of searchable terms has been made available in the home page search box (see Figure 1).

	ma		Cerca
	agenti mascheranti	^	
	Anleihemarkt		
	Anti-Mafia-Staatsanwalt		
	Anti-Mafia-Staatsanwaltschaft		
	antimafia		
Bartala Terminologica Ted	Antimafia-Bezirksstaatsanwaltschaf		
Portale terminologico tede	attacco infor ma tico		
GEROM è una piattaforma lessicale e te	st cavo sotto ma rino		volume "Terminologia e corpora per il
dinamico, pensata per supportare l'anali	si controllo ma ncato		menti tradizionali e nuove applicazioni"
ad ambiti quali la politica, la società e la	ci crimine contro l'u ma nità		no al progetto e inserito nella collana "La
tedesca e italiana.	Direzione distrettuale antimafia		a, traduzione e interpretazione".
GEROM coniuga una serie di repertori le	Direzione investigativa antimafia	L' CICILLO AUGIOITIALO OCILO	, accolte testuali e lessicali inserite nel portale
la consultazione di corpora testuali creat	i ad hoc per la realizzazione dei	è disponibile qui.	
singoli glossari. Fare clic su Domini per v	isualizzare l'elenco completo	44 magnis 0045 . La varri	
dei temi.		nubblico alla facoltà UNIN	T di Roma, Informazioni

Figure 1: Home page search box

2.2 Response page

The next functional block of the CGI program is designed to produce the response pages to the clients by extracting all the required information from both the terminological knowledge-base and the corpora. To carry out this complex task, the first step must consist in printing the header lines which are required to create a UTF-8 encoded HTML page. In addition, all the Javascript and CSS codes which have been edited by Claudio Fantinuoli to enrich the web page with graphical features are included in this section, too.

After the header block declaration has been set, the next step regards selecting and retrieving text samples from the files which are stored in the web server. To do this, the program iteratively scans all the lines the terminological file is made of to find any possible match with the user's query string. All pattern matching tests are managed by regular expressions, which provide the possibility to define exact or partial matching rules. In our case, any pattern matching test will pass only if the user input string and a specific field of the terminological file are proven to be identical. After a correct match is found, the whole set of metadata, flexional alternatives and homonyms tied to the query string are saved in dedicated variables and an array is set to store the links to the corpora which are eligible to contain the text patterns to find.

2.3 Corpus-driven approach

After collecting all the useful data from the terminological reference file, the program uses the same pattern matching technique to sample a limited number of lines from the subject-related corpora. This step is very important because it allows the system to provide users with a short list of examples which represent the basic stage to understand the real use of language in its own specific context. Yet, differently from the operations involving the terminological dataset, the lookup and text extraction activity made on corpora is driven by both the query string and its own related homonyms and flexional variants, too.

Summing up, the whole text mining process implements two nested loops. The former manages the pattern matching between the input term and all the lines the terminological file is made of; the latter matches the content of the selected corpora with both the term to search and, if any, its semantic and morphologic variants. After concluding the pattern matching activity on one corpus, the same processing is run on the corpus which contains the corresponding terminological information in the other language GEROM offers. During the above mentioned tasks, pertinent linguistic data is retrieved, collected and properly organized in an html page that will be transmitted to the user web browser as the end of the terminological file is met (see Figure 2).

© Center for Promoting Ideas, USA www.ijllnet.com

an p	al a fermine on annetterstrungtruß					
comat	fia (info - Browse)	For	ti esterne Externe	Quellen		
	Umweltverschmutzung	dan	no ambientais in Lie	onue Done	Trace and	
DE	Umweltschaden	Unit	veitverschmutzung i	in Lingues - Po	275	
	Umweitgefährdung					
	danno ambientale					
т	degrado ambientale					
	inquinamento ambientale ChudWenger &					
)eu	tsch	Italiano				
finitio	n: Belastung, Schädigung der natürlichen Umwelt durch Schmutz, schädliche Stoffe	Definizione: È dan Indiretto di una mor	no ambientale quals	siasi deleriorar	nento significativo e misu da questutima	rabile, diret
uelle:	Duden	Fonte: Enciclopedia	a Treccani	ning asserundin	roa quest orona.	
Kontes Quelle Kookk	t Die Umwöhreischmutzung im Todes dreisch soll jetzt gestoppt werden. taz trezz, withem, lifegal, Folgen der U., Auswirtungen der U., kampt gegen die U. eren	in Contesto for Fonte II F	realtà, tra metà anni no mbato con "evidente in rganizzazione nei terre ratto Quotidiano	ovanta e l'inizio d estimabile danni ni tra Caserta e f	el 2000, il patturne di ogni ge o ambientale" e ingente gua: Vapoli.	nere veniva tagno per
Kontes Quelle Kookk Komm	t Die Umweltverschmutzung im Todesdreisch" soll jetzt gestoppt werden. taz ertenz: ertem, Hegal, Folgen der U., Auswirkungen der U., Kampf gegen die U. enzig ert	in Contesto ta Fo Fonte II F Cooccorrenze nik	realtà, bia metà anni no mbato con "evidente in rganizzazione nel ferre ratto Guolediano avante, evidente, inesti	ovanta e l'inizio d estimabile danni ni tra Caserta e f mabile, duraturo	el 2000, il patturne di ogni ge o ambientale" e ingente guar Napoli, 	nere veniva tagno per
Kontex Quelle Kookk Komm POS <u>?</u>	t Die Umweltverschmutzung im Todesdreisch" soll jett gestoppt werden. taz ertenz: erten, Hogal, Folgen der U., Auswirsungen der U., Kampf gegen die U. ertar NOLRH	In Contesto for Fonte II F Cooccorrenze nik Commento	realtà, tra metà anni no mbato con "evidente in rganizzazione nel terrei fatto Guolidiano evante, evidente, inestri	ovanta e l'inizio d estimabile danni ni tra Caserta e f mabile, duraturo	el 2000, il patturne di ogni ge o arribientale" e ingente guai Vapoli. . danno all'arribiente	nëre veniva tagno ser
Kontes Quelle Kookk Komm POS <u>?</u> Korpu	t Die Umwehverschmutzung im Todesdreisch" soll jett gestoppt werden. taz ertnez: ertnem, Hogal, Folgen der U., Auswirnungen der U., Kampf gegen die U. ertar ROUN1 & Erweiterte Ansicht	Contesto in the form of the form of the form of the former	realt), tra metà anni n mbato con "evidente in rganizzazione nel terrei rato Guolidiano avante, evidente, inesti DUN m	ovanta e Tinizio d estimabile danni ni tra Caserta e f mabile, duraturo	el 2000, il patturne di ogni ga o ambientale" e ingente guat Napoli 	nere ventva tagno per
Kontex Quelle Kookk Komm POS ? Korpu	t Die Umwehverschmutzung im Todesdreieck" soll jett gestoppt werden. taz ertnez: ertnem, liegal, Folgen der U., Auswirnungen der U., Kampf gegen die U. ertar NOUNT & Envelterte Anscht	Contesto in In To Fonte II P Concorrenze min Commento POS <u>?</u> Ni Corpus Avanzat	realt), tra metà anni ni mbato con "evidente in galaczazione nel terrer atto Guolidiano ivante, evidente, inesti 2004 m. 0	ovanta e finizio d estimabile danni ni tra Caserta e f mabile, duraturo	el 2009, il patturne di opri ge o ambientale" e ingente guar (apoli, 	nere veniva Jagno per
Kontex Quelle Kookk Komm POS ? Korpu	Die Umwelheischmutzung im Todes dreisch soll jetzt gestoppt werden. taz ertenz, erbern, lifegal, Folgen der U., Auswirkungen der U., Kampf gegen die U. estar roQUN1 s Erweitlerte Ansicht verprestet. Das Gros der Umweltverschmutzung aber geschieht mit ensor an	Contesto in to Fonte II F Cooccorrenze riti Commento POS 2 N/ Corpus Avanzad	realtà, tra metà anni ni mbato con "evidente in rganizzazione nel terrer rato Guolodiano evante, evidente, inesti DUN m 0	ovanta e finizio d estimabile danni ni tra Caserta e f mabile, duraturo	el 2000, il patturne di ogni ga o ambientale" e ingente guar l'apoli danno all'ambiente	nere veniva tagno per
Konter Quelle Kookk Komm POS ? Korpu	t Die Umwelheischmutzung im Todesdreich" ablijetd gestoppt werden. taz ertenz, ertem, liegal, Folgen der U., Auswirkungen der U., Kampf gepen die U. estar NOUNT is Erweiterte Anscht werpestet Das Gros der Umweltverschmutzung aufgen ibl gener der Umweltverschmutzung aufgeneinen umweltverschmutzung in Kalabrien weltertnis ungelo	In Contesto to	reati), tra metà anni no mbato con "widente in reganizzazione nel terre- rato Quolediano vvante, evidente, nesti DUN m D resentano anzi un	ovanta e finizio d estimabile danni ni tra Caserta e t mabile, duraturo danno	el 2000, il patturne di opri ge o ambientale" e ingente guar Aspoli danno all'ambiente ritervante: basta com	nere veniva tagno per siderate
Kontes Quelle Kookk Komm POS ? Korpu Q	t Die Umweltweischmutzung im Todes dreisch zoll jetzt gestoppt werden. taz ertreaz, ertrem, liegal, Folgen der U., Auswirkungen der U., Kampf gepen die U. estar NOUNI * & Erweitente Ansicht verpestet. Das Grös der Umweltverschmutzung aber geschieht mit eigen ist jener der Umweltverschmutzung in Kalabienen alligemeinen Auswirkungen entimet Umweltverschmutzung auf den menschlichen	Contesto fin To Fonte II Cooccores POS 7 NI Corpus Avanzat	realit), tra metà anni no moto con "woderte in riganizzazione nel terre l'atto Guueldiano vivante, evidente, inesti JUN m D unesentano anzo un vo la consecuenza	ovanta e finizio d estimabile danni ni tra Caserta e f mabile, duraturo, danno ambientale	el 2000, il patturne di opri ge ambientale" e ingente guar Vapoli. .danno all'ambiente rilevante: basta com la s	nere veniva tagno per
Kontex Quelle Kookk Komm POS ? C	t Die Umwelheirschmutzung im Todesdreisch soll jett gestoppt werden. taz ertreaz, erthem, liegal, Folgen der U., Auswirkungen der U., Kampf gepen die U. estar NOUNF : s Erweitenfe Ansicht verpestel. Das Gros der Umweltverschmutzung aber geschieht mit egen ist gener der Umweltverschmutzung in Kalabrein weiterbin aligemeinen Umweltverschmutzung auf den menschlichen : als und mögliche Folgen. Umweltverschmutzung auf den menschlichen : als und mögliche Folgen. Umweltverschmutzung auf den menschlichen :	Contesto in	realit), tra metà anni no mbato con "widerte in riganizzazione ne terre l'atto Guolediano viratte, evidente, inesti DUN m o resentano anzo un on la conseguenza che il	ovanta e finicio d estimabile danni ni tra Caserta e f mabile, duraturo danno ambientale danno ambientale	el 2000, il patturne di opri ge ambientale" e ingente guar Ageoti danno all'ambiente ritevante: basta com la s è transitato da Ceng Giuglia	nere ventva tagno per siderate to a
Kontes Quelle Kookk Komm POS ? Korpu Q Q Q	t Die Umwelheischmutzung im Todesdreisch ablijktigestopptweiden. taz errenz, eiterm, liegal, Folgen der U., Auswirnungen der U., Kampt gegen die U. erstar nOUNT te Erweterte Ansicht werpestet. Das Gros der Umweltverschmutzung aber geschieht mit im Kalabrien weitertin ungelo Auswirkungen erterme Umweltverschmutzung auf den meschlichen aus und mögliche Fölgen umweltverschmutzung sternweiter Null vertramt. Die Umweltverschmutzung umweltverschmutzung einer sternweise Null vertramt. Die Umweltverschmutzung im "Todesdreiseck" soll jet	In Contesto In	reatili, tra metà anni no mbato con "evidente in instato con "evidente in rato Guoldiano Duthi m D mesentano ano un on la conseguenza che il ui mai, per fortuna. il	ovanta e Tinisio d estimabile dannin in tra Caserta e 1 mabile. duraturo danno ambientale danno ambientale	el 2000, il patturne di opi ge ambientale" e ingente guar dapoli .danno all'ambiente ritevante: baista com la s è transitato da Ceng Giuglia si rivela tale. In alter	nativa co
Kontex Quelle Kookk Komm POS ? Korpu O O O O	d Die Umweltweischmutzung im Todes dreisch" soll jedt gestoppt werden. tac tac tac retrenz, in eigen, Folgen der U., Auswirkungen der U., Kampf gegen die U. estar rockhit te Erweiterte Ansicht werpestet. Das Gros der Umweltwerschmutzung eigen sit gener der eigen sit gener der eigen sit gener der eigen sit gener der Auswirkungen extreme Umweltverschmutzung auf den meschlichen als und mögliche Folgen umweltverschmutzung einer ber Umweltverschmutzung einer umweltwerschmutzung einer einer umweltwerschmutzung einer umweltwerschmutzung	Contesto in To Fonte Coccorrenter in Commento POS 7 Ni Corpus Avanzat O unitate e rapp Corpus Avanzat O unitate e rapp Contesto quanto quando quan	reatil, tra metà anni no mbato con "evidente in mbato con "evidente in ratto Quolostano varatte, evidente, inesti DUN m o nesentano anzi un on la conseguenza che il u mai, per fortuna. il dente inestimabile	ovanta e Tinisio d estimabile dannin ini tra Caserta e t mabile, duraturo danno ambientale danno ambientale danno ambientale danno	el 2000, il patturne di opi ge ambientale" e ingente guar dapoli. .danno all'ambiente ritevante: basta con la s è transitato da Ceng Giuglia si nvela tale. In alter " e ingente guadager Vorg	nere veniva tagno per siderate po a nativa co io per
Kontei Quelle Kookk Komm POS ? Korpu O O O O O O O Vertei	t Die Umweltweischmutzung im Todes dreisch" soll jetd gestoppt werden. taz wertenz, in depat, Folgen der U., Auswirnungen der U., Kampf gepen der U. estar nOUNT te Erweterte Ansicht werpestet. Das Gros der Umweltwerschmutzung aber geschieht mit eaglingen ist jener der auswirkungen eiterme Umweltwerschmutzung im Kalabnen weiterhin aus und mögliche Fölgen umweltwerschmutzung im "Todesdreisch" soll jett ber Umweltwerschmutzung im "Todesdreisch" soll jett umweltwerschmutzung im "Todesdreisch" soll jett ber Umweltwerschmutzung im "Todesdreisch" soll jett tung im Körpus	Contesto for	reatil, tra metà anni ni mbato con 'widente in reganizzacione nel terre riganizzacione nel terre viante, evidente, inesti 2004 m o ni la conseguenza che il u mai, per fontuna, il dente inestimabile ete di risarcimento	ovanta e hnisio d estimabile dann ni ta Caseta e t mabile, duraturo danno ambientale danno ambientale danno ambientale danno ambientale danno	el 2000, il patturne di opri ge a ambientale" e ingente guar dapoli. .danno all'ambiente .fRevante: basta com la s e transitato da Ceng Giuglia si riveia tale. In alter " e ingente guadage " ceare una rete di	nere veniva tagno per siderate po a nativa co io per aziende e
Kontei Guelle Kookk Komm POS ? Korpu Q Q Q Q Q Q Vertei	t Die Umweltweischmutzung im Todes dreisch" soll jett gestoppt werden. taz wernez, ertem, liegal, Folgen der UL, Auswirkungen der UL, Kampf gegen die U. NOUNT & Erweiterte Anscht werpestet Das Gros der Umweltwerschmutzung aber geschieft mit. eigen ist jener der Umweltwerschmutzung in Kalabienen als und mögliche Folgen dir Umweltwerschmutzung einftatt und die songespatie seenweise Mull verbrannt. Die Umweltverschmutzung im "Todesdreisch" soll jett	Contesto for for Fonte III Cooccorrenze in Commento POS 7 M/ Corpus Avanzat Corpus Avanzat Corpus Avanzat Corpus Avanzat Congue Avanz	reatil, tra metà anni ni mbate con 'widente in impanizzacione nel terre atto Gueldano invante, evidente, inesti 2004 m o resentano ano un on la conseguenza che il u mai, per fortuna, il dente inestimable ete di risarcimento dei	ovanta e hinsio d estimabile danni ini ta Caserta e t mabile, duraturo danno ambientale danno ambientale danno ambientale danno ambientale	el 2000, il patturne di opri ge a ambientale" e ingente guar tapoli. .danno all'ambiente rilevante: basta com la s fi ranstato da Ceng Giuglia si rivela tale. In alter " e ingente guadage riorg ; creare una rete di sog	nere veniva tagno per siderate po a nativa co io per aziende e
Kontei Guelle Kookk Korpu O O O O O Vertei	t Die Umweltweischmutzung im Todes dreisch" ablij jett gestoppt werden. Taz Trazen and tenn, liegal, Folgen der UL, kanauf gegen der	In Contesto in Sin Contesto in	reatil, tra metà anni ni mbato con "evidente in regarizzazione nel tetra ganizzazione nel tetra arto Gueletano avante, evidente, inesti DUN m o uresentano anzo un resentano anzo un on la conseguenza che il u mak, per fortuna, il dente inestimabile ete di risarcimento del	ovanta e finisio d estimabile danni ini ta Caserta e t mabile, duraturo ambientale danno ambientale danno ambientale danno ambientale danno ambientale danno ambientale	el 2000. Il patturne di opri ge ambientale" e ingente guar Vapoli. .danno all'ambiente rileivante: basta com la s è transitato da Ceng Giuglia si rivela tale. In alter " e ingente guadapr Vorg : creare una rete di sog	nere veniva tagno per siderate po a nativa co io per azende e
Kontei Guelle Kookk Komm POS ? Korpu O O O O O O O Vertei	d Die Umweltweischmutzung im Todes dreisch 1 obi jetd gestoppt werden. ta: rater in einen (liegal, Folgen der U., Auswirkungen der U., Kampf gegen die U. ratorn i ratorn i	Contesto in Contesto in Concorrenze Nos 2 No Corpus Avancat Corpus Avancat Quando quar concorrenze ioni concr Distribuzione n	reatil, tra metà anni no mbato con "evidente in riganizzazione nel terrer altro Quelviano avarite, evidente, inesti DUN m o resentano ano un on la coherte, inesti o no na coherte in gi anal, per fortuna. il dente inestimabile ete di risarcimento del ete corpus	ovanta e Inicis d estimable danne en la Caserta e I mable, duraturo danno ambientale danno ambientale danno ambientale danno ambientale danno ambientale	el 2000, il patturne di opi ge ambientale" e ingente guar dapoli. .danno all'ambiente ritevante: basta con la s è transitato da Ceng Giuglia si rivela tale. In alter " e ingente guadapr rorg ; creare una rete di sog	nere veniva tagno per siderate po a nativa co io per aziende e
Kontei Quelle Kookk Komm POS ? Korpu Q Q Q Q Q Vertei	t Die Umwelhwischmutzung im Todes dreisch" soll jetzt gestoppt werden. taz wertenz, in tehm, liegal, Folgen de U., Auswirkungen der U., Kampt gegen die U. soluert s Erweiterte Anscht werpestet. Das Gros der Umweltverschmutzung aber geschieht mit eigen ist jener der Jumweltverschmutzung im Kalabnen weiterin aus und mögliche Folgen umweltverschmutzung erhtmut die sener ssenweise Null verbrann. Die Umweltverschmutzung im "Todesdreisch" soll jet Musetverschmutzung im "Todesdreisch" soll jet für Gres Spiegel 4.667 % 2.33.3 %	Contesto fon	reatil, tra metà anni ni matric con 'widente in inganitzacione nel terri ganitzacione nel terri dio Gueldano ni la conseguenza che il ui mai, per fontuna il dente inestimabile ete di risancimento del ete corpus	ovanta e finicio di estimabile danni ini ta Caserta e 1 mabile, duraturo danno ambientale danno ambientale danno ambientale danno ambientale danno ambientale	el 2000, il patturne di opel ge a ambientale" e ingente guar dapoli. .danno all'ambiente ritevante: baista com la s è transitato da Ceng Giuglia si rivela tale. In alter " e ingente guadage Vorg ; creare una rete di sog.	nere veniva tagno per siderate po a nativa co io per aziende e
Kontei Quelle Kookk Komm POS ? Corpu O O O O O O Vertei	d Die Umweltweischmutzung im Todes dreisch" ablijetd gestoppt werden. Taz werden in legal. Folgen der UL Auswirkungen der UL Aampf gegen der UL abligen in der Begen der Umweltweischmutzung ihr Kalabrien werdenten der in tal gener der Umweltweischmutzung ihr Kalabrien werdenten abligen in der Greisen Umweltweischmutzung ihr Kalabrien werdenten abligen indigener Folgen umweltweischmutzung ihr Kalabrien werdenten abligen der Greisen Umweltweischmutzung ihr Kalabrien werdenten ber Umweltweischmutzung ihr "Todesdreisech" soll jet Umweltweischmutzung ihr, Todesdreisech" soll jet kar der Spegel 4 667 % faz 2 33.3 %	In Contesto for for for for the concentration of the Concentration of the Concentration of the Contest of the C	reată, tra metă anni ni mbate con "widente in inganizzacione nei terre atră Guzieldano invante, evidente, inestă 2004 m o in la conseguenza che ii u mai, per fortuna, ii dente inestimabile ete di risarcimento dei et corpus	ovanta e finisio d estimabile danne en la Caserta e ? mabile. duraturo danno ambientale danno ambientale danno ambientale danno ambientale mero risuitati/	el 2000, il pattirrie di opi ge o ambientale" e ingente guar tapoli. .danno all'ambiente rifeviante: basta com la s è transitato da Ceng Giuglia si riveia tale. In alter " e ingente guadage r'org ; creare una rete di sog	nere ventva tagno zer siderate po a nativa co o per aziende e
Kontes Quelle Kookk Komm POS 7 O O O O O O O O O O O O O O O O O O O	t Die Umweltweischmutzung im Todes dreisch" abt jett gestoppt werden. Tag terrenz in term, liegal, Folgen de UL, Auswirkungen der UL, Kampf gegen de UL. terrenz in terrenz im Schweiter Schmutzung in der geschieft mit degen ist joner der Umweltweischmutzung aber geschieft mit degen ist joner der Umweltweischmutzung auf den menschischen all sund mögliche Folgen umweltweischmutzung auf den menschischen de und mögliche Folgen umweltweischmutzung auf den menschischen all sund mögliche Folgen umweltweischmutzung auf den menschischen auf umweltweischmutzung auf den menschischen auf umweltweischmutzung auf den menschischen de unseitweiser Mutzung auf den menschischen auf umweltweischmutzung auf den menschischen den einer Umweltweischmutzung auf den den schischen den einer Umweltweischmutzung auf den menschischen auf umweltweischmutzung auf "Todesdreisech" soll jett tung im Korpus	In Contesto in the test of	reatil, tra metà anni ni matar con "evidente in matar con "evidente in riganizzazione nel tener aria Gualettano avante, evidente, inesti DUN m o uresentano anzo un on la conseguenza che il u mak, per fortuna, el conte in estimabile ete di risarcimento del et corpus	ovanta e Inicis d estimable danne en la Caserta e I mable, duraturo ambientale danno danno	el 2000. Il patturne di opri ge ambientale" e ingente guar Vapoli. .danno all'ambiente rileivante: baista com la s è transtato da Ceng Giuglia si rivela tale. In alter " e ingente guadapr torg : creare una rete di sog	nere verha tagno ger siderate jo à nativa co o per aziende e



Another important feature offered by GEROM regards the possibility to overcome the reduced number of elements the basic response page is designed to return and enhance the visualization of linguistic information by providing users with a wider range of concordance lines extracted from the corpora in use. To achieve this goal, a specific link to a dedicated CGI script has been set in each tab the user response page holds. For the script be able to work, relevant parameters like the name of the corpus where information should be retrieved from and the term to find - the original query string or one of its linguistic variants - must be transmitted. Like before, also in this case a HTML page is returned to the user, but it will be opened in a new window for readability reasons (see Figure 3). As for the text lines which are sampled from the corpora to be included into the response page, it is worth noting that the adoption of a dedicated hypertext markup structure was mandatory to display the whole data in a KWIC (Key Word In Context) compliant format.

GEROM

	Target ¹ Global search Context Left Right Source	Search text	Regex2 Smart search ³ Image: Image of the search of the searc		
Show 1	0 ∽ entries			Search:	
Let	it 🔶	Key 🔶	Right	÷	Source 🕴
0	tume di ogni genere veniva tombato con "evidente inestimabile	danno ambientale	" e ingente guadagno per l'organizzazione nei terreni tra Ca		Il Fatto Quotidiano
	della basilitza dell'Assa di Oscala sera la construcción de ll	danno ambientale	è transitato da Cengio a Giugliano". Inquinamento della fald		ll Fatto Quotidiano
	dalla ponífica dell'Acha di Cenglo, con la conseguenza che li				
	dalla bonifica dell'Acha di Cengio, con la conseguenza che il se lo scalpore per certe notizie ha vita breve, al contrario i fra un omicidio e l'altro [*] . Così Carmine Schiavone, ex boss dei cas	danni ambientali salesi e collaboratore	dello sversamento di rifiuti tossici nelle cave della Campania e di giustizia dal 1993. Ma se lo scalpore per certe notizie ha vita bi	reve, al con	Il Fatto Quotidiano trario i danni
0	dalia boninica dell'Acha di Cengio, con la conseguenza che il se lo scalpore per certe notizie ha vita breve, al contrario i fra un omicidio e l'altro". Così Carmine Schiavone, ex boss dei cas ambientali dello sversamento di rifiuti tossici nelle cave della Camp luoghi uilibrio dell'ecosistema" guando guasi mai, per fortuna, il	danni ambientali salesi e collaboratore pania hanno una dur danno ambientale	dello sversamento di rifiuti tossici nelle cave della Campania e di giustizia dal 1993. Ma se lo scalpore per certe notizie ha vita bi rata lunghissima. Fino a quando non vengono rimossi. E qui casca si rivela tale. In alternativa come un evento dannoso il cui ri	reve, al coni l'asino. "lo	Il Fatto Quotidiano trario i danni certe cose, come i Il Fatto Quotidiano
	dalla bonifica dell'Acha di Cengio, con la conseguenza che il se lo scalpore per certe notizie ha vita breve, al contrario i fra un omicidio e l'altro". Così Carmine Schiavone, ex boss dei cas ambientali dello sversamento di rifiuti tossici nelle cave della Cam luoghi ullibrio dell'ecosistema" quando quasi mai, per fortuna, il rischio di tumori, definire azioni concrete di risarcimento del	danni ambientali salesi e collaboratore opania hanno una dur danno ambientale danno ambientale	dello sversamento di rifiuti tossici nelle cave della Campania e di giustizia dal 1993. Ma se lo scalpore per certe notizie ha vita bi rata lunghissima. Fino a quando non vengono rimossi. E qui casca si rivela tale. In alternativa come un evento dannoso il cui ri ; creare una rete di aziende e soggetti che promuovano e difen	reve, al coni l'asino. "lo	Il Fatto Quotidiano trario i danni certe cose, come i Il Fatto Quotidiano La Repubblica
	dalla boninica dell'Acha di Cengio, con la conseguenza che il se lo scalpore per certe notizie ha vita breve, al contrario i fra un omicidio e l'altro". Così Carmine Schiavone, ex boss dei cas ambientali dello sversamento di rifiuti tossici nelle cave della Camp luoghi uilibrio dell'ecosistema" quando quasi mai, per fortuna, il rischio di tumori, definire azioni concrete di risarcimento del rattore. Ma non sono di alcuna utilità e rappresentano anzi un	danni ambientali salesi e collaboratore pania hanno una dur danno ambientale danno ambientale danno ambientale	dello sversamento di rifiuti tossici nelle cave della Campania e di giustizia dal 1993. Ma se lo scalpore per certe notizie ha vita bi rata lunghissima. Fino a quando non vengono rimossi. E qui casca si rivela tale. In alternativa come un evento dannoso il cui ri ; creare una rete di aziende e soggetti che promuovano e difen rilevante: basta considerate la superficie che occupano, che no	reve, al coni l'asino. "lo	Il Fatto Quotidiano trario i danni certe cose, come i Il Fatto Quotidiano La Repubblica Corriere della Sera
	dalia boninica dell'Acha di Cengio, con la conseguenza che il se lo scalpore per certe notizie ha vita breve, al contrario i fra un omicidio e l'altro". Così Carmine Schiavone, ex boss dei cas ambientali dello sversamento di rifiuti tossici nelle cave della Cam luoghi uilibrio dell'ecosistema" quando quasi mai, per fortuna, il rischio di tumori; definire azioni concrete di risarcimento del rattore. Ma non sono di alcuna utilità e rappresentano anzi un aperto degli scarti industriali di mezzo Paese. Incalcolabili i	danni ambientali salesi e collaboratore pania hanno una dur danno ambientale danno ambientale danno ambientale danno ambientale danni ambientali	dello sversamento di rifiuti tossici nelle cave della Campania e di giustizia dal 1993. Ma se lo scalpore per certe notizie ha vita bi rata lunghissima. Fino a quando non vengono rimossi. E qui casca si rivela tale. In alternativa come un evento dannoso il cui ri ; creare una rete di aziende e soggetti che promuovano e difen rilevante: basta considerate la superficie che occupano, che no e sanitari, soprattutto nel triangolo della Terra dei Fuochi,	reve, al con l'asino. "Io	Il Fatto Quotidiano trario i danni certe cose, come i Il Fatto Quotidiano La Repubblica Corriere della Sera Il Fatto Quotidiano
0 0 0 0 0 0	dalia boninica dell'Acha di Cengio, con la conseguenza che il se lo scalpore per certe notizie ha vita breve, al contrario i fra un omicidio e l'altro". Così Carmine Schiavone, ex boss dei car ambientali dello sversamento di rifiuti tossici nelle cave della Cam luoghi uilibrio dell'ecosistema" quando quasi mai, per fortuna, il rischio di tumori, definire azioni concrete di risarcimento del rattore. Ma non sono di alcuna utilità e rappresentano anzi un aperto degli scarti industriali di mezzo Paese. Incalcolabili i uppo allora guidato da Raul Gardini centinaia di miliardi per i	danni ambientali salesi e collaboratore pania hanno una dur danno ambientale danno ambientale danno ambientale danni ambientali danni ambientali	dello sversamento di rifiuti tossici nelle cave della Campania e di giustizia dal 1993. Ma se lo scalpore per certe notizie ha vita bi rata lunghissima. Fino a quando non vengono rimossi. E qui casca si rivela tale. In alternativa come un evento dannoso il cui ri ; creare una rete di aziende e soggetti che promuovano e difen rilevante: basta considerate la superficie che occupano, che no e sanitari, soprattutto nel triangolo della Terra dei Fuochi, provocati dall' impianto di Cengio. Si va avanti così, fra in	reve, al con l'asino. "lo	Il Fatto Quotidiano trario i danni certe cose, come i Il Fatto Quotidiano La Repubblica Corriere della Sera Il Fatto Quotidiano Corriere della Sera

0	onomica. Quando si parla di "Ecomafia" si fa riferimento ai danni ambientali	, che si pagano in termini di sicurezza, malattie e morti, e ch		I Fatto	Quotic	liano
Showing 1 to	10 of 11 entries		Previous	1	2	Next

Figure 3: Enhanced response page

3. Future development

Although the basic framework the GEROM web platform is founded on has achieved a stable operating structure, for some aspects the project should be still considered as a work-in-progress. In particular, further development are planned to increase the linguistic resources offered by the system in terms of both subject-specific corpora and external references users may get access to. Gerom is freely accessible at the following web address: http://gerom.unint.eu/it/index.html

References

- Zhang, C.; Wu, D. (2012). Bilingual Terminology Extraction Using Multi-level Termhood. In: The Electronic *Library*, 30(2), pp. 295-309
- Fantinuoli, C.; Proia, F. (eds.). (2015). Terminologia e corpora per il dialogo italo-tedesco. Strumenti tradizionali e nuove applicazioni. Roma, Aracne.
- Robbins, A. (2015). Gawk: Effective AWK Programming, 4th Edition. Universal Text Processing and Pattern Matching. Sebastopol, O'Reilly Media.
- Scott, G.; Shishir, G.; Gunther, B. (1999). CGI Programming with Perl, 2nd Edition. Sebastopol, O'Reilly Media.

Jeffrey, E.F.F. (2006). Mastering Regular Expressions, 3rd Edition. Sebastopol, O'Reilly Media.

- Mari-Carmen Marcos et al. (2006). Usability evaluation of online terminology databases. In: Hipertext.net, 4. Accessed at: http://www.hipertext.net [19/10/2015]
- Zengin, B. & Kacar, I.G. (2011). Turkish EFL Academicians' Problems Concerning Translation Activities and Practices, Attitudes towards the Use of Online and Printed Translation Tools, and Suggestions for Quality Translation Practice. In: Turkish Online Journal of Educational Technology, 10(2), 274-286.
- Ahmet Aker, Monica Lestari Paramita, Marcis Pinnis (Tilde), Robert Gaizauskas. 2014. Bilingual dictionaries for all EU languages. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 2839-2845.
- Tatiana Gornostay and Andrejs Vasiljevs. 2014. Cloud Terminology Services Facilitate Specialised Lexicography Work. In Proceedings of the XVI EURALEX International Congress: The User in Focus, 15-19 July 2014, Bolzano/Bozen, pp. 621-629.
- Vasiljevs, A., Pinnis, M., & Gornostay, T. (2014). Service model for semi-automatic generation of multilingual terminology resources. In: Terminology and Knowledge Engineering 2014. Accessed at: https://hal.archives-ouvertes.fr/hal-01005874 [19/10/2015]