

What languages does the internet speak?

2007 statistics on the presence of Latin languages on the web

In 1992, English took up almost all of the web. In 15 years, the situation has changed considerably: the presence of English has fallen by half and the Latin languages are increasingly present. This study, the result of cooperation between the Latin Union, an intergovernmental organization, and the NGO Networks and Development Foundation (Funredes), enjoys the support of the Délégation générale à la langue française et aux langues de France.

Presentation of the study

The Latin Union and Funredes have been studying the place of Latin languages on the web since 1996. This is the seventh study, and it has been modernised to include some new features. In addition to Spanish, French, Italian, Portuguese, Romanian and German, the languages now studied include Catalan. It still contains a compilation of everything that has been measured since the very first study, but it focuses on the comparative study of the presence of Latin languages in relation to English (and German) on websites and in blogs, a new area of measure. The study also presents detailed results for each linguistic area, which enables a comparison of web page productivity per country.

Methodology

The methodological principle of the study has remained unchanged, that is, the rigorous selection of equivalent terms in all the languages studied, the measurement of the number of pages indexed in a search engine corresponding to each of these terms in the different languages studied and, finally, the statistical construction and validation of the results. The methodology has had to be slightly revised however in order to adapt to developments in search engines and to offset the effects of the invasion of advertising on the internet, which has tended to distort the results.

Evolution of languages on the web between 1998 and 2007

Web pages in	1998	2000	2001	2003	2005	2007
ENGLISH	75 %	60 %	52 %	49 %	45 %	45 %
SPANISH	2.53 %	4.79 %	5.50 %	5.31 %	4.60 %	3.80 %
FRENCH	2.81 %	4.18 %	4.45 %	4.32 %	4.95 %	4.41 %
PORTUGUESE	0.82 %	2.25 %	2.55 %	2.23 %	1.87 %	1.39 %
ITALIAN	1.50 %	2.62 %	3.08 %	2.59 %	3.05 %	2.66 %
ROMANIAN	0.15 %	0.21 %	0.18 %	0.11 %	0.17 %	0.28 %
GERMAN	3.75 %	2.85 %	6.75 %	6.80 %	6.94 %	5.90 %
CATALAN	-	-	-	-	-	0.14 %
OTHERS	3.44 %	22.20 %	23.68 %	19.65 %	33.43 %	36.54 %

The statistics show that Latin languages have stood still in relation to English. This state of affairs is indicative of a phenomenon concerning the space indexed by each search engine. In 2003, English grew in the indexed space and this situation seems to have recurred over the period 2005-2007. The trend does not denote a decline in the presence of Latin languages in cyberspace, but rather a decline in their presence in the spaces indexed by the main search engines, in particular as a result of the significant development of the Chinese language on the internet. The only real increase has been for Romanian, which must be a result of its recent entry into the European Union.

Productivity of internet users by language

By dividing the percentage of web pages by language by the number of internet users by language, we obtained a productivity index of web pages by language. These figures show clearly the low productivity of Latin countries in relation to French-speaking and English-speaking countries; only Italian shows a satisfactory capacity for web page production in the national language.

	P = Pages	I = Internet users	P/I*	
			2005	2007
ENGLISH	45.00 %	31.71 %	1.26	1.42
SPANISH	3.80 %	8.79 %	0.53	0.43
FRENCH	4.41 %	5.06 %	1.21	0.87
ITALIAN	2.66 %	2.72 %	0.83	0.98
PORTUGUESE	1.39 %	4.10 %	0.63	0.34
ROMANIAN	0.28 %	0.43 %	0.31	0.66
GERMAN	5.90 %	5.10 %	1.04	1.16
CATALAN	0.14 %	0.18 %	x	0.74

** a productivity index of 1 is a "normal" result, less than 1, a "poor" result and more than 1, a "considerable" result.*

**The study may be consulted in full at: http://dtil.unilat.org/LI/2007/index_fr.htm
Information: dtil@unilat.org**