

***El Departamento de Estadística del ITAM***

anuncia la siguiente sesión de

***EL SEMINARIO ALEATORIO***

que con el título

***THE WEB ARCHIVES: A TIME-  
MACHINE IN YOUR POCKET!***

será impartida por

***A.S.A. Roehrl***

Swiss Federal Institute of Technology, Lausanne, Switzerland

***RESUMEN***

Taking an interdisciplinary approach, we discuss both the technical issues of creating archives of the World Wide Web (as suggested at [www.archive.org](http://www.archive.org)), and the possible socio-political relevance of such archives in the future. As the Internet becomes the Ever and Everywherenet, the Web archives may become a memory of mankind, a sort of time-machine to go back into the past. We present the hardware and software concepts, and an initial analysis, of a highly scalable and extendable approach to archive a fully queryable copy of the ever-changing Web. The purpose is not to compete with the efforts at [www.archives.org](http://www.archives.org), but to present research results that may be useful in any future archiving project of the Web. The software approach is unique in that the search strategy of the Web crawler is based on capture-recapture techniques (from statistics), rather than the common brute-force method of scanning as many Web pages as possible. This includes estimates on the quality of the scans. This is a difficult challenge, due to the rapidly changing character of the Web. The Swiss-Tx supercomputer is used as the hardware platform of choice, because of its inconnect switch, which allows to change the connections between the different nodes on the fly to enhance the performance of the parallel search database. Finally, we show that the constitution of coherent Web archives will necessarily be of interest to researchers of all specialities, since the memory of human kind can be the best heritage to be left to future generations.

***Fecha: Viernes 17 de marzo***

***Hora: 11:00***

***Lugar: Salón 307***

*El Seminario Aleatorio está destinado tanto a profesores como a estudiantes, por lo que el Departamento de Estadística agradece a los profesores que colaboren invitando a sus alumnos a estas sesiones.*