



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería de Sistemas

**Especificación de un sistema de aprendizaje de
patrones de navegación web utilizando gramáticas
libres de contexto probabilísticas de hipertexto**

TESIS

Para optar el Título Profesional de Ingeniero de Sistemas

AUTOR

Augusto Parcemon CORTEZ VÁSQUEZ

ASESOR

Dra. Luzmila Elisa PRO CONCEPCIÓN

Lima, Perú

2016

Referencia bibliográfica

Cortez, A. (2016). *Especificación de un sistema de aprendizaje de patrones de navegación web utilizando gramáticas libres de contexto probabilísticas de hipertexto*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Sistemas]. Repositorio institucional Cybertesis UNMSM.



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Académico Profesional de Ingeniería de Sistemas

Acta de Sustentación de Tesis

Siendo las *16:51* horas del día *30* de Junio del año 2016, se reunieron los docentes designados como miembros de Jurado de la Tesis, presidido por el Ing. Gilberto Anibal Salinas Azaña (Presidente), el Lic. Jorge Luis Chávez Soto (Miembro) y la Dra. Luzmila Elisa Pro Concepción (Miembro Asesor) para la sustentación de la Tesis Intitulada: **“ESPECIFICACIÓN DE UN SISTEMA DE APRENDIZAJE DE PATRONES DE NAVEGACIÓN WEB UTILIZANDO GRAMÁTICAS LIBRES DE CONTEXTO PROBABILÍSTICAS DE HIPERTEXTO”**. Por el Bachiller: CORTEZ VÁSQUEZ, AUGUSTO PARCEMON; para optar el Título Profesional de Ingeniero de Sistemas.

Acto seguido de la exposición de la Tesis, el presidente invitó al graduado a dar las respuestas a las preguntas establecidas por los Miembros del Jurado.

EL graduado en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez a las observaciones y preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, el graduado obtuvo la nota de *17* (En letras) *diecisiete*.

A continuación el Presidente de Jurados el Ing. Gilberto Anibal Salinas Azaña, declara al graduado **Ingeniero de Sistemas**.

Siendo las *18:00* Horas, se levantó la sesión.

.....
Presidente
Ing. Gilberto Anibal Salinas Azaña

.....
Miembro
Lic. Jorge Luis Chávez Soto

.....
Asesor
Dra. Luzmila Elisa Pro Concepción

Universidad Nacional Mayor de San Marcos

Universidad del Perú, DECANA DE AMÉRICA

Facultad de Ingeniería de Sistemas e Informática

Facultad de Ingeniería de Sistemas e Informática

Título

Especificación de un Sistema de Aprendizaje de patrones de navegación web utilizando gramáticas libres de contexto probabilistas de hipertexto

Tesis, para optar el título profesional de Ingeniero de Sistemas

Autor : CORTEZ VASQUÉZ, Augusto Parcemón

Asesor : PRO CONCEPCION, Luzmila Elisa

Fecha: Junio del 2016

RESUMEN

Uno de los problemas que reviste real interés en el área de minería de datos de uso de la web es capturar las actividades de los usuarios durante su conexión y extraer patrones de comportamiento que permitan definir sus preferencias con el fin de mejorar el diseño de futuras paginas adaptando las interfaces de los sitios web a los usuarios individuales. En esta investigación se pretende ofrecer en primer lugar una presentación de los fundamentos metodológicos del uso de lenguajes probabilistas para identificar sitios web más relevantes o visitados. En segundo lugar se representa las sesiones web mediante grafos y gramáticas libres de contexto probabilísticas, de tal forma que las sesiones que tengan mayor probabilidad son consideradas las más visitadas o más preferidas, por tanto las más relevantes en relación a un tópico determinado. Se pretende la especificación de una herramienta para procesamiento de sesiones web obtenidas a partir de log de servidor representado mediante gramáticas probabilistas libres de contexto.

Palabras claves: Hipertexto, gramáticas probabilística, patrones de navegación, aprendizaje de patrones de navegación, navegación hipertextual.

National University of San Marcos

(University of Peru, Dean of America)

Faculty of Engineering and Information Systems

Academic Professional School of Systems Engineering

Title

Specifying a Learning System web browsing patterns using probabilistic context-free grammars hypertext

Thesis to choose the professional title of Systems Engineer

Author: CORTEZ VASQUÉZ, Augusto Parcemón

Advisory: PRO CONCEPCION, Luzmila Elisa

Date: Junio del 2016

SUMMARY

One issue that is of real interest in the area of data mining Web usage is to capture user activities during connection and extract patterns of behavior that help to define your preferences in order to improve the design of future pages It interfaces adapting websites to individual users., This research aims to provide first a presentation of the methodological foundations of the use of probabilistic languages to identify relevant or websites visited. Secondly Web sessions by graphs and probabilistic context-free grammars, so that sessions that are most likely represented are considered the most popular or most preferred, therefore the most important in relation to a particular topic. Specifying a tool for processing Web meetings obtained from server log represented by probabilistic free grammars of context is intended.

Keywords: Hypertext, probabilistic grammars, browsing patterns, learning navigation patterns, hypertext navigation