

FICHA DE RESULTADOS DE PROYECTOS

1. DATOS DEL PROYECTO

Título: Visualización del Sistema de Transporte Urbano de Gijón (VISTUG)

Investigador/a responsable (perteneciente al IUTA): Ignacio Díaz Blanco

Tfno: 985182663

E-mail: idiaz@uniovi.es

Otros investigadores:

Abel A. Cuadrado Vega

Antonio Robles Álvarez

Francisco Javier García

Fernández Daniel Pérez López

Becario:

Gonzalo Moreno Ávila

Empresas o instituciones colaboradoras:

(Indíquese el municipio donde están ubicadas las empresas y/o instituciones, su grado de interés y de colaboración y en qué ha consistido dicha colaboración):

El proyecto recibió el visto bueno de EMTUSA (principal beneficiario potencial) para su desarrollo.

2. BREVE MEMORIA DESCRIPTIVA DEL PROYECTO

Resumen ejecutivo:

El proyecto plantea el desarrollo de una aplicación de técnicas de analítica visual, basadas en la combinación de análisis de datos y visualización de datos para extraer conocimiento útil del estado de la red de autobuses municipal de Gijón a partir de los datos abiertos puestos a disposición por el Ayuntamiento de Gijón en <http://datos.gijon.es>. La idea del proyecto, cuyo principal beneficiario potencial podría ser la empresa municipal de transportes urbanos, parte de la apreciación de la calidad y detalle de la información disponible en los servicios de <http://datos.gijon.es>, que incluyen información geolocalizada de todos los autobuses en tiempo real, así como de las paradas y trayectos. La disponibilidad de información espacial y temporal de todos los autobuses, líneas y paradas sugiere el potencial de estos datos (más allá de su representación directa en una aplicación web o en un dispositivo móvil) para extraer conocimiento, tanto de la calidad de servicio como de los factores que influyen en la misma. Este proyecto tiene como objetivos desarrollar una aplicación que permita supervisar el sistema de transporte urbano de Gijón, y explorar las posibilidades de aplicación de técnicas de visualización y de análisis de datos para presentar la información de forma novedosa, en la que el uso de visualizaciones alternativas, en combinación con algoritmos de análisis de los datos permita descubrir patrones o modelos de comportamiento en el servicio de transportes.

Los objetivos se han cubierto satisfactoriamente a través del desarrollo de una aplicación interactiva que permite visualizar, para un periodo concreto (a partir de datos extraídos previamente de datos.gijon.es) los retrasos en los autobuses respecto a las previsiones mostradas en el sistema de información (que se expone en las marquesinas). La visualización permite analizar la información en el espacio y en el tiempo, para las distintas líneas y trayectos lo que permite potencialmente el análisis de las causas de los retrasos al facilitar la identificación de estos con otra información espacial, temporal (ej. atascos, horas punta) o del sistema (características de las líneas, y trayectos) disponible al usuario.

Las tareas del proyecto se desarrollaron en las fases previstas: 1. Estudios preliminares. 2. Extracción de datos. 3. Prototipado de algoritmos de análisis inteligente. 4. Programación de los algoritmos online. 5. Programación de la interfaz. 6. Puesta a punto y mejora.

El proyecto cubre holgadamente los objetivos planteados en el proyecto y su uso ha permitido además detectar algunos patrones interesantes en el comportamiento de la red de autobuses (mostrando líneas u horas concretas con mas retrasos). Más allá de los objetivos planteados, se ha iniciado el desarrollo de una segunda aplicación web de visualización de las velocidades por horas, trayectos y líneas que utiliza D3.js, una librería novedosa para visualización de datos basada en javascript. Esta aplicación se encuentra actualmente en fase de desarrollo.

Objetivos inicialmente planteados:

De acuerdo con la memoria del proyecto los objetivos eran:

*Este proyecto tiene como **objetivos** desarrollar una aplicación que permita supervisar el sistema de transporte urbano de Gijón, y explorar las posibilidades de aplicación de técnicas de visualización y de análisis de datos inteligentes (IDA, intelligent data analysis) para presentar la información de forma novedosa, en la que el uso de visualizaciones alternativas, en combinación con algoritmos de análisis de los datos (generando, por ejemplo, índices de calidad del servicio, o detectando deficiencias en el mismo) permita descubrir patrones o modelos de comportamiento en el servicio de transportes.*

Grado de consecución de los objetivos planteados:

Los objetivos se han cubierto satisfactoriamente a través del desarrollo de dos aplicaciones interactivas de visualización de información del servicio de autobuses:

Visualización interactiva de retrasos en los autobuses Se ha desarrollado una aplicación interactiva en Processing (un lenguaje desarrollado en el MIT, para desarrollo de interfaces de visualización de datos interactivas, basado en java) que, a partir de la información disponible en la web del Ayuntamiento de Gijón datos.gijón.es permite una visualización interactiva de los retrasos en los autobuses, respecto a las predicciones proporcionadas en el sistema de información de las marquesinas. Esta aplicación cubre holgadamente los objetivos planteados en el proyecto y su uso ha permitido además detectar algunos patrones interesantes en el comportamiento de la red de autobuses (mostrando líneas u horas concretas con mas retrasos).

Aplicación web interactiva de visualización de velocidades en trayectos (en desarrollo). Aunque como se indica la aplicación anterior cubre los objetivos, de cara a valorar otra tecnología de visualización de datos vía web, se inició el desarrollo de una aplicación web interactiva para la visualización de las velocidades de los autobuses para los distintos trayectos y líneas, hora a hora, utilizando la librería D3.js, basada en javascript, que permite el desarrollo de aplicaciones de visualización interactivas en web. En el momento de escribir este informe, la parte visual está muy desarrollada y se está trabajando en la estimación de las velocidades medias por horas en puntos espaciales concretos (las marquesinas), para lo que es necesario el uso de algoritmos de interpolación (redes neuronales), aunque hay estimaciones preliminares que permiten dar una idea de la aplicación.

La primera de las aplicaciones en las que se ha trabajado cubren satisfactoriamente los objetivos planteados en el proyecto y la segunda, aún en fase experimental, supone una aportación adicional al proyecto al explorar una tecnología del estado del arte en visualización de datos (D3.js) así como la adquisición de conocimiento en otro problema interesante, como es la visualización del patrón de

velocidades de la red de autobuses en la ciudad. Es importante mencionar que el desarrollo del proyecto ha tenido un grado de complejidad elevado en la parte de tratamiento de los datos, debido al carácter asíncrono y deslocalizado de los datos disponibles en el servidor <http://datos.gijon.es> (los datos están disponibles en intervalos de tiempo irregulares, para un número variable de autobuses en cada actualización). Ello, con la dificultad añadida del carácter multivía del problema (análisis en varias vías o dimensiones: horas/marquesinas/líneas/trayectos) requirió una considerable inversión de tiempo en la elección de las estructuras para almacenamiento de la información, y tratamiento de los datos hasta obtener un archivo estandarizado, con datos horarios por marquesina/líneas/trayecto, listo para utilizar en la aplicación de visualización. Otro problema encontrado fue la no disponibilidad de datos en el servidor de datos.gijon.es, que en algunos casos interrumpe el servicio. La gestión de esos datos faltantes supuso una dificultad adicional.

Tareas realizadas (indicando claramente la participación del personal becario en ellas):

Para la consecución de los resultados descritos en el punto anterior, se han desarrollado las siguientes tareas:

- **Estudios preliminares.** El becario realizó un estudio detallado de los datos y la documentación disponible en los servicios web de datos.gijón.es, incluyendo estructura y campos disponibles de los datos, formatos de las llamadas a los servicios, etc. Asimismo, hizo un estudio del estado de la técnica.
- **Extracción de datos.** Partiendo de pruebas de concepto iniciales que se le suministraron, el becario desarrolló varias aplicaciones en Python para la adquisición y tratamiento de la información proporcionada por el servidor datos.gijon.es cuyo cometido fue la obtención, organización de los datos, filtrado y eliminación de datos erróneos y normalización, para la generación de archivos CSV con una estructura predeterminada (formato matriz) que contienen toda la información visualizada por la aplicación interactiva.
- **Análisis de datos.** También mediante scripts de Python, se realizaron los cálculos de los retrasos para la visualización de retrasos mediante comparación entre los tiempos reales de llegada a cada marquesina con los tiempos previstos.
- **Interfaz de visualización de datos.** El becario desarrolló una aplicación interactiva en Processing (un lenguaje desarrollado en el MIT, para la creación de interfaces de visualización de datos interactivas, basado en java) que, a partir de la información de archivos CSV obtenidos en la fase de extracción de datos, permite una visualización interactiva de los retrasos en los autobuses, respecto a las predicciones proporcionadas en el sistema de información de las marquesinas. El sistema tiene las siguientes características:
 - Permite la visualización en un mapa de los retrasos, mediante una librería de mapas navegable (unfolding processing) sobre el que se representan los retrasos/adelantos mediante puntos de tamaño y color variable con el retraso en cada marquesina
 - Dispone de un sistema de *timeline* que permite al usuario desplazarse mediante una barra o "slider", hora a hora en el tiempo, para ver la evolución de los retrasos. La selección del intervalo de tiempo mostrado puede hacerse también mediante un calendario desplegable.
 - Permite además filtrar los retrasos/adelantos por minutos (por ej. marcar todas las marquesinas con un retraso de 1')
 - Finalmente, permite mostrar información de contexto por cada marquesina, indicando los retrasos y adelantos en las líneas correspondientes a esa marquesina.
- **Evaluación y comprobación.** En paralelo con las fases anteriores, el becario hizo un importante esfuerzo de evaluación y comprobación de los resultados. Esta tarea fue difícil por la gran cantidad de situaciones singulares (por ejemplo, la lista variable de autobuses en funcionamiento, ausencia de datos de algunos autobuses en periodos largos, puestas a cero de las predicciones por parte del servicio, etc.) que tuvieron que ir siendo subsanadas a lo largo del desarrollo del proyecto.

Resultados obtenidos:

Los principales resultados son el desarrollo de dos aplicaciones interactivas:

- Aplicación de escritorio para la visualización interactiva de retrasos en los autobuses programada en lenguaje Processing.
- Prototipo preliminar de aplicación web para la visualización interactiva de velocidades en los autobuses programada en lenguaje Processing.

Divulgación de los resultados: deben indicarse publicaciones, ponencias en congresos, trabajos fin de carrera, patentes, jornadas divulgativas, etc.

Los resultados obtenidos se han presentado recientemente en las jornadas de seguimiento de proyectos organizadas por el IUTA en diciembre de 2013.

Trabajos futuros:

Como se indicó anteriormente, la visualización espacio-temporal de velocidades de los autobuses es un problema complejo, ya que se parte de datos asíncronos (a intervalos irregulares de tiempo y en posiciones también irregulares) de posiciones GPS de los autobuses de los que hay que obtener las velocidades en instantes (horas) y lugares concretos (trayectos entre marquesinas). En el momento de escribir este informe, la parte visual está bastante avanzada y se está trabajando en la estimación de las velocidades medias por horas en puntos espaciales concretos (trayectos entre marquesinas), para lo que es necesario el uso de algoritmos de interpolación (redes neuronales).

Se plantea en un futuro inmediato investigar el uso de distintos tipos de redes neuronales del estado del arte (*radial basis functions* RBF, *support vector machines* SVM, etc.) para la estimación de las velocidades en posiciones e intervalos temporales regulares.

Asimismo una línea de futuro trabajo que será esencial es la valoración del potencial de las herramientas de visualización novedosas aquí empleadas (como D3.js) en problemas de análisis de datos abiertos de cara a nuevos desarrollos.

3. MEMORIA ECONÓMICA

Financiación		Personal	Inventariable	Fungible	Otros gastos
IUTA	SV-13-GIJÓN-1	3000 €	-----	-----	-----
Otras fuentes	Referencia proyectos/contratos				
Personal Becario	Nombre	Gonzalo Moreno Avila			
	Tareas	1. Estudios preliminares. 2. Extracción de datos. 3. Prototipado de algoritmos de análisis inteligente. 4. Programación de los algoritmos online. 5. Programación de la interfaz. 6. Puesta a punto y mejora.			
	Período	Desde septiembre 2013 hasta diciembre de 2013			