# Práctica obligatoria de Integración de Sistemas Informáticos

# Curso 2018-2019

# Master Ingeniería Informática

## Descripción

El objetivo del caso práctico es trabajar en un escenario de integración de datos e integración funcional en el dominio de las ciudades y las necesidades de sus ciudadanos. Se trata de aprovechar los datos en abierto para proporcionar servicios y aplicaciones para favorecer distintos aspectos de la vida de los ciudadanos y ayudar al gobierno de la ciudad. Un ejemplo es la planificación óptima de los desplazamientos en relación con la movilidad y transporte en las ciudades, un campo en que se necesita desarrollar planes eficientes de movilidad urbana fomentando el transporte sostenible, desplazamiento a pie, la bicicleta o el uso del transporte público colectivo para reducir el tráfico y la consiguiente contaminación. Para ello es posible cruzar datos sobre tráfico tanto privado como público, alquiler y uso de bicicletas (por ejemplo, BiciMAD <https://www.bicimad.com/>), uso compartido de transporte, etc. así como de observatorio de ciudadanos (por ejemplo, empleando redes sociales, blogs, foros en los que los ciudadanos opinan y reflejan quejas sobre diferentes servicios).

Además de la movilidad y transporte existen otros dominios como la seguridad (seguridad vial, seguridad ciudadana, convivencia, ...), emergencias (prevención y extinción de incendios y salvamentos), comercio (ubicaciones, actividad comercial, etc.), educación, sociedad y bienestar, deportes y salud, entre otros. El portal de datos abiertos del Ayuntamiento de Madrid (<https://datos.madrid.es/portal/site/egob/>) proporciona conjuntos de datos en estos dominios, así como APIs para su integración en diversas ventanas temporales. Otras ciudades disponen de portales similares, lo que favorecería la integración y obtención de comparativas y análisis en distintas zonas geográficas.

Otras fuentes de datos, a modo de ejemplo, que pueden integrarse:

* Portales inmobiliarios para alquiler y venta de viviendas en distintas ciudades (<https://www.idealista.com/>).
* Portales web sobre transporte público y privado que contienen información de itinerarios, infraestructuras, etc. (Metro de Madrid <https://www.metromadrid.es/>).
* Portales sobre medioambiente y calidad de vida
* Portales sobre salud, educación, deporte, etc.
* Mapas, planos, e información geográfica (Google Maps, Carto DB, ...)
* Otros

En relación con los observatorios ciudadanos que pueden integrarse con las fuentes de datos anteriormente mencionadas, en el caso de la ciudad de Madrid en la categoría “Sociedad y Bienestar” existe el conjunto de datos de “Avisos ciudadanos” con incidencias en la vía pública transmitidas al ayuntamiento a través de distintos canales, así como datos de participación ciudadana a través de debates y propuestas. Además, Twitter y otras redes permitirían monitorizar la vida de la ciudad:

1. ¿de qué se está hablando en la ciudad? obteniendo temas sobre los que se habla en la ciudad a partir de Twitter (los temas pueden estar definidos a priori o no).
2. ¿qué está pasando en la ciudad? Obteniendo alertas sobre distintos tipos de incidencias (accidentes, atascos, manifestaciones, emergencias) según posiciones o zonas.
3. ¿qué sentimientos hay en el ambiente? Qué sentimientos expresan los ciudadanos sobre determinados temas (positivos, negativos o neutros).

En la asignatura de Integración de Sistemas se estudian tecnologías que facilitan la integración de aplicaciones a distintos niveles: datos, funcional o lógico y presentación. El objetivo de la práctica de este año consiste en desarrollar un esquema de integración de datos, diseñando una arquitectura software y hardware e implementar una prueba de concepto para facilitar el acceso a los datos, y su posterior análisis y visualización. Evidentemente, se trata de aplicar los conocimientos que se irán adquiriendo durante la asignatura, de manera que no será necesario implementar nuevos métodos o algoritmos para el tratamiento de esta información sino que será necesario, únicamente, integrar sistemas ya existentes.

Así, la solución a desarrollar en la práctica deberá integrar información y funcionalidad de distintas fuentes, almacenándola y proporcionando algún mecanismo de visualización. El desarrollo será un demostrador que no tiene por qué coincidir con el diseño propuesto, podría ser un subconjunto (por ejemplo, con menos número de fuentes de datos o menor funcionalidad a integrar).

A continuación, se proporciona exclusivamente a modo de ejemplo un escenario de integración, consistente en un servicio de planificación y ayuda al uso de la bicicleta en la ciudad. El objetivo sería desarrollar un servicio que ayude al Integración de datos sobre tráfico en la ciudad según histórico e información en tiempo real para proponer zonas en las que es adecuado el uso de la bicicleta. Se podría integrar información de (1) los itinerarios recomendados, (2) mapas de disponibilidad para alquiler de bicicletas, (3) alertas de eventos en la ciudad que dificultarían el uso de la bicicleta (eventos programados, emergencias obtenidas en tiempo real, etc.), (4) datos históricos sobre accidentes con bicicletas para ver la siniestralidad según puntos geográficos y franjas horarias, (5) transporte de bici en otros medios para completar rutas según requisitos del usuario, (6) condiciones meteorológicas y otros recursos que fueran de utilidad para el ciudadano. Como APIs a integrar (1) planificadores de rutas en la ciudad, (2) software para análisis del lenguaje natural de tweets, etc.

Se pide:

1. Definir el valor añadido que proporciona al usuario el escenario de integración (conocimiento que no es posible obtener de las fuentes de datos por separado).
2. Seleccionar, recopilar y analizar las fuentes a integrar comprobando el modelo de datos y el formato, el histórico y periodo contemplado, la periodicidad de actualización, etc.
3. Diseñar un sistema software que recoja la información de las fuentes seleccionadas en el punto anterior (en tiempo real o fuentes off-line) basándose en los conceptos de integración vistos en la asignatura.
4. Elegir un sistema de almacenamiento adecuado y almacenar la información obtenida. La tecnología propuesta para el almacenamiento de la información recogida, que debe ajustarse a un paradigma de arquitectura distribuida sobre la que aplicar el paradigma map-reduce.
5. Diseñar un sistema que basándose en los datos almacenados realice los análisis requeridos y que además visualice la información utilizando distintos formatos (líneas temporales, mapas, gráficos estadísticos, etc.). El análisis de los datos se realizará a través de un sistema externo (sea una API, una librería software externa, etc.). Este diseño debe basarse en alguno de los paquetes de software estudiados en la asignatura.
6. Plantear las consultas básicas que proporcionen la información requerida por el usuario. Por ejemplo, en el escenario proporcionado a modo de ejemplo, algunas consultan podrían ser (a) cuál es la ruta con menos tráfico para hacer en bici entre Principe Pío y Cuatro Caminos en el momento de la consulta y (b) mostrar en una línea temporal de 2 años la siniestralidad de accidentes de bicicleta en las calles de la zona centro de Madrid y Barcelona.
7. Para la visualización pueden utilizarse herramientas como librerías de visualización en javascript ([protovis](http://mbostock.github.com/protovis/), [Google API visualization](http://code.google.com/intl/es-ES/apis/chart/interactive/docs/gallery.html), [dygraphs](http://dygraphs.com/)),  Gephi (para visualización de grafos, <http://gephi.org/>) o software [Tableau](http://www.tableau.com/) entre otras.
8. Implementar una prueba de concepto del sistema diseñado. Esta prueba no tiene que incluir toda la propuesta teórica, puede ser una parte. Como requisito para la la prueba de concepto se deberán integrar al menos **tres** fuentes de datos y desde el punto de vista de la integración funcional será necesario integrar al menos **dos** servicios de terceros, por ejemplo, para analizar el texto de los comentarios de los ciudadanos en redes sociales. Tanto la integración de datos como la funcional se hará de acuerdo a las tecnologías estudiadas en la asignatura. La funcionalidad de esta prueba de concepto deberá validarse con los profesores de la asignatura.

## Normas generales de la práctica.

La práctica se empezará a desarrollar tan pronto como se tengan conocimientos básicos de la asignatura. El desarrollo se hará de forma progresiva, para lo cual se dividirá en distintas fases, acordes con los contenidos aprendidos en las clases. Estas fases van a permitir un avance gradual de acuerdo a lo enseñado en el curso. Como se indica en el cronograma de la asignatura publicado en AG, el **5 de Noviembre se llevará a cabo la validación por parte de los profesores de la asignatura de los diseños planteados por los alumnos**. Cualquier duda podrá ser consultada bien durante los días de prácticas en aula informática, bien a través del foro de la asignatura en Aula Global. En este último caso el profesor de prácticas intentará resolverla lo antes posible (por favor, no enviar correos con dudas concretas sobre la práctica, siempre consultar por medio del foro de Aula Global). Durante el curso, si es necesario, se publicará material adicional que facilite la resolución de estas dudas y que ayude en las determinadas fases del desarrollo.

## Grupos

La práctica se realizará por parejas. Los alumnos deben ser capaces de argumentar y defender cualquiera de las decisiones que se hayan tomado en la elaboración del caso práctico, demostrando así su participación real en la realización del caso práctico (recordad que hacerle la práctica al compañero no es un ejemplo de compañerismo). La prueba de concepto será evaluada durante una sesión de defensa a la que ambos alumnos deberán asistir y responder a las preguntas que realice el profesor. **Los grupos de prácticas deberán estar formados y deberán ser notificados al profesor de prácticas antes del jueves 8 de octubre de 2017**.

## Estructura de la memoria

La memoria para cada fase de la práctica deberá contener los siguientes apartados:

* Portada: Con el título del trabajo y los datos de los miembros del grupo.
* Resumen ejecutivo: Una página con una panorámica del trabajo realizado que debería describir el objetivo, un párrafo describiendo la solución diseñada, aspectos clave y conclusiones. Este resumen debe contener todo lo necesario para familiarizarnos con el contenido de la memoria final.
* Análisis y diseño del problema de integración propuesto: Esta sección describirá las interfaces empleadas, el objetivo de su integración, es decir, la funcionalidad que proporcionará, la arquitectura software diseñada. Habrá que identificar los componentes software que hacen falta, las interfaces que existen entre ellos y la forma en que se proporcionarán dichas interfaces. Para la arquitectura hardware basta tener en cuenta las indicaciones mencionadas más arriba.
* Prueba de concepto: Este apartado describirá la prueba de concepto desarrollada incluyendo instrucciones para la ejecución de la prueba de concepto en la máquina virtual presentada.
* Conclusiones: Este último apartado destacará los objetivos alcanzados, los problemas encontrados durante el desarrollo de la práctica y las lecciones aprendidas.

## Entregas

Se realizará una única entrega. Se espera que cada grupo de prácticas entregue una memoria que contemple los apartados descritos en la sección anterior y la máquina virtual en la que se implemente la prueba de concepto de integración realizada o código fuente según sea el caso.

• **Fecha de la evaluación de la prueba de concepto:** **10 de diciembre de 2017** (último día de clase). Cada grupo ejecutara en presencia de los profesores la prueba de concepto y responderá a las cuestiones planteadas por los profesores. Para la entrega de la memoria de la práctica y la prueba de concepto implementada se comunicará el día con la suficiente antelación.

• La calificación obtenida será un 90 % de la nota final del curso.

Las entregas se realizarán en aula global a través de una tarea. Solamente uno de los dos alumnos que forman parte del grupo deberá realizar la entrega online. Cada entrega deberá consistir en un fichero PDF cuyo nombre debe seguir el siguiente formato: NIA1\_NIA2.pdf, NIA1, NIA2 los NIAS de los alumnos que forman el grupo. Además, será necesario ofrecer una implementación de la solución de integración propuesta, que se entregará mediante una máquina virtual ejecutable en el entorno Oracle VM VirtualBox. Conviene asegurarse de que esta máquina virtual se ejecuta sin problema en distintas máquinas.

## Criterios de evaluación

Los aspectos que se tendrán en cuenta a la hora de evaluar el trabajo y su peso son:

* Complejidad del problema de integración propuesto: 40%. Se valorará la búsqueda de fuentes a integrar y la creatividad de la solución propuesta
* Calidad de la memoria: 20%
* Adecuación de la solución de integración: 40%