

Arxiu documental de la Societat del Gran Teatre del Liceu

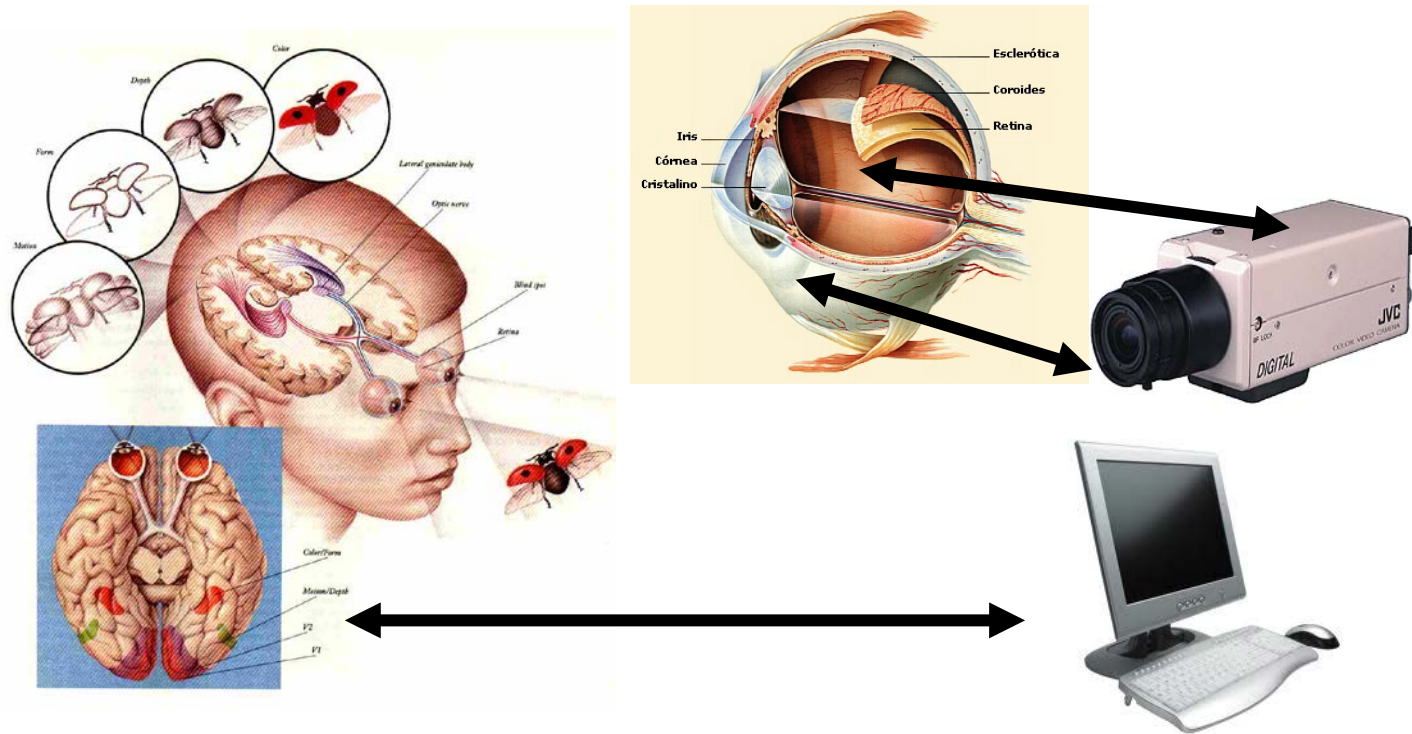
Recerca en anàlisi de documents

Alicia Fornés, Josep Lladós
Centre de Visió per Computador - UAB



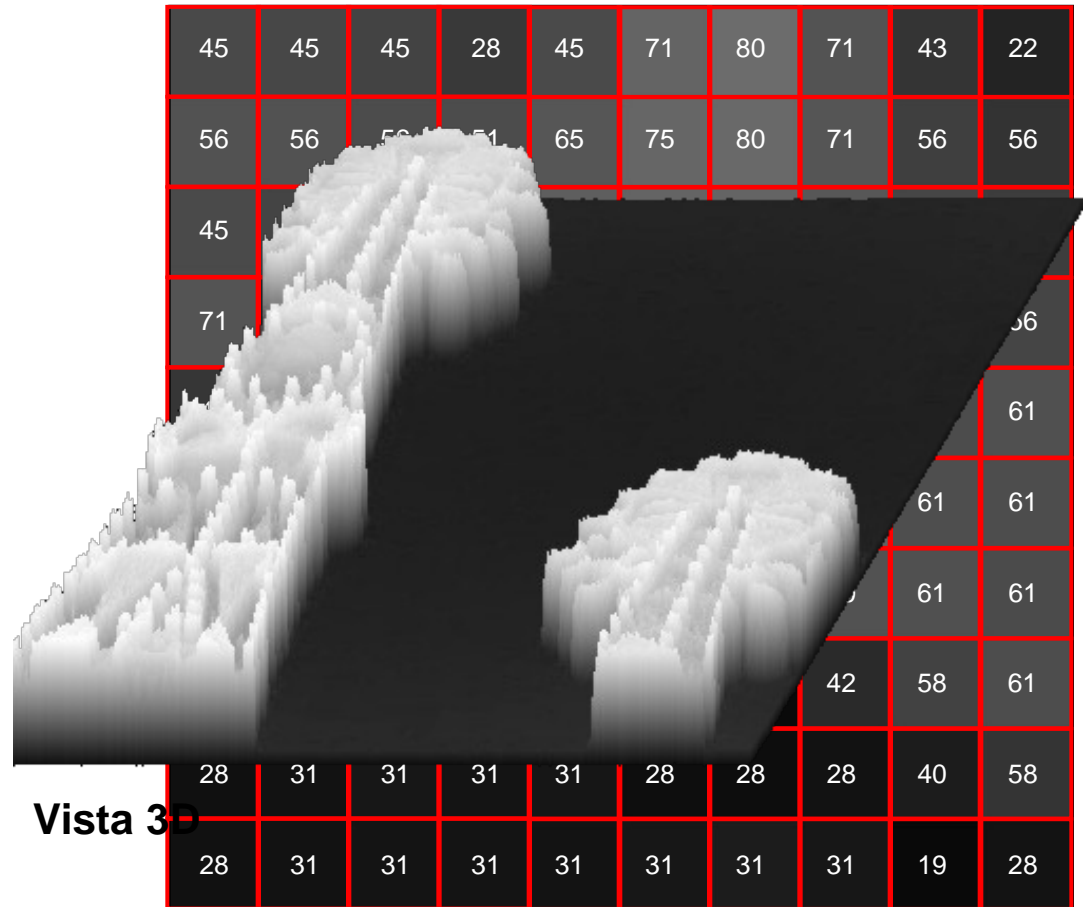
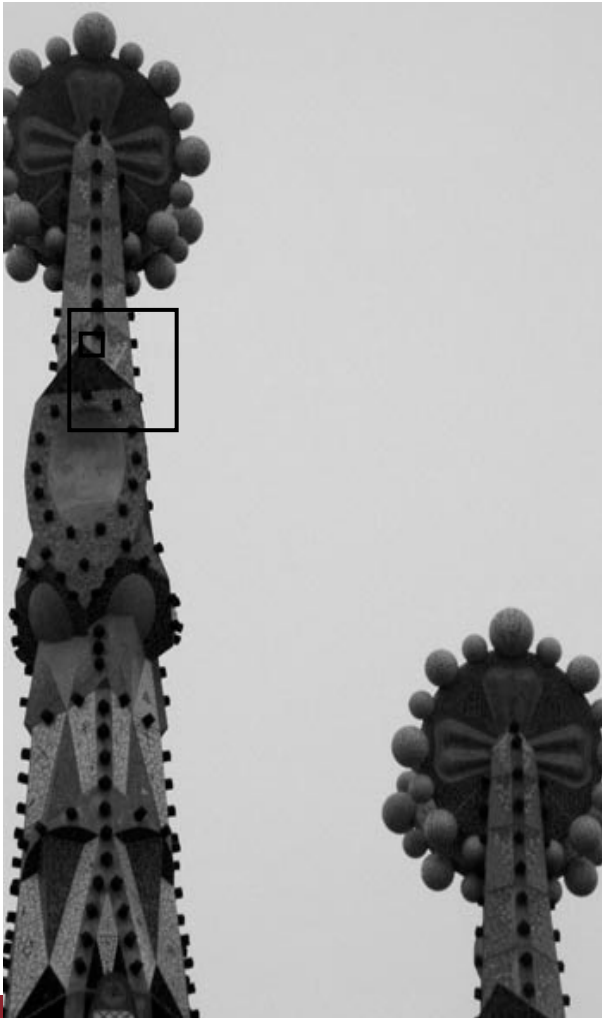
La Visió per Computador

La Visió per Computador és la disciplina de la informàtica que fa que les màquines hi vegin.



Què és una imatge digital?

Una imatge digital és una funció bidimensional $I(x,y)$ tal que cada punt x,y (*pixel*) té el valor de la intensitat de la llum en aquest punt.

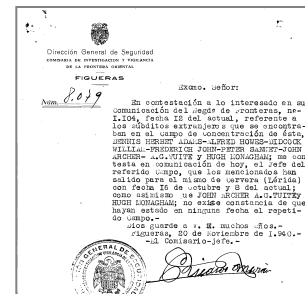


Vista 3D

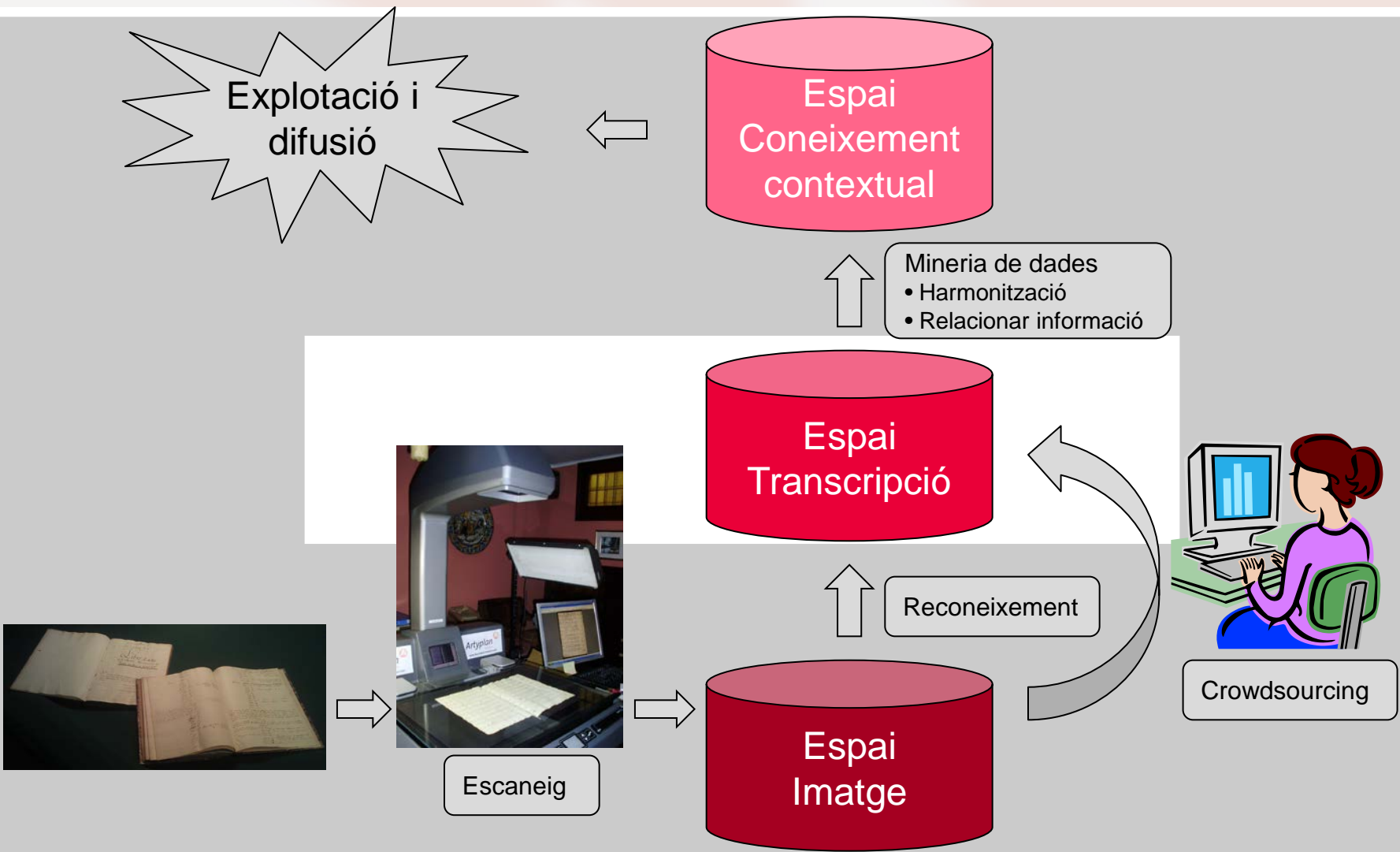
Imatge en nivells de gris

L'Anàlisi de Documents

- Les col·leccions de documents històrics emmagatzemats en arxius i biblioteques són un bé públic singular, que forma la memòria col·lectiva i l'evolució de les nostres societats.
- Les campanyes de digitalització massiva garanteixen la preservació en forma digital però no l'accés als continguts i la seva valorització.
- **L'anàlisi i reconeixement d'imatges de documents històrics** és l'àrea de visió per computador que aborda el problema de reconèixer de forma automàtica el contingut del document (text imprès o escrit a mà, o elements gràfics).
- Les plataformes de crowdsourcing són sistemes basats en la web, on milers de voluntaris transcriuen el contingut del document.



Arquitectura



Fases del projecte

1. Fase de digitalització i catalogació
2. Fase d'extracció d'informació
3. Aplicacions

Fase 1: Digitalització i Catalogació

1. Escaneig dels documents
2. Catalogació dels documents
3. Creació de pàgina web que habiliti cerques

Fase 2.1: Extracció d'informació textual

- Reconeixement textual (transcripció)

- Diari del conserge
- Diari de secretaria
- Llibres d'Actes
- Contractes
- Programes de ma
- Factures
- Certificats

- Habilitació de cerques

- Cercar informació sobre una persona
- Cercar relacions entre persones (descobrir informació)



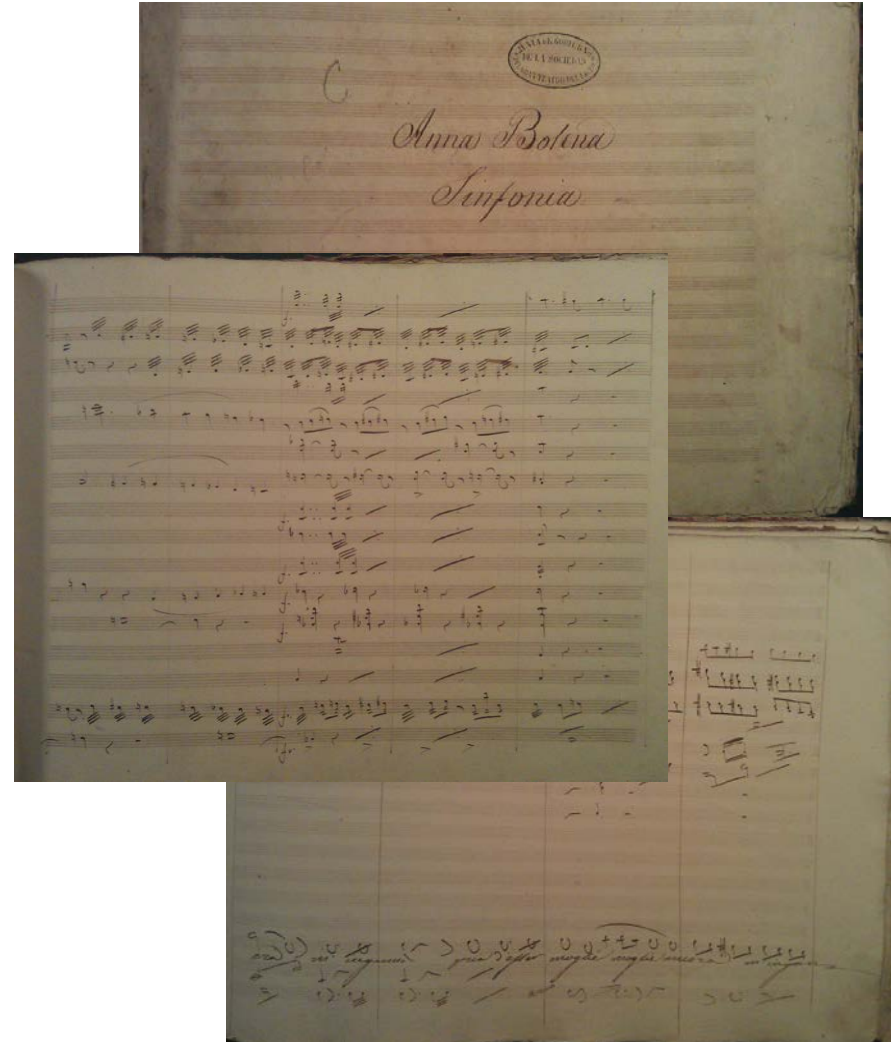
Fase 2.2: Extracció d'informació gràfica (1)

- Partitures Musicals

- Editar el llibre
- Generar l'audio



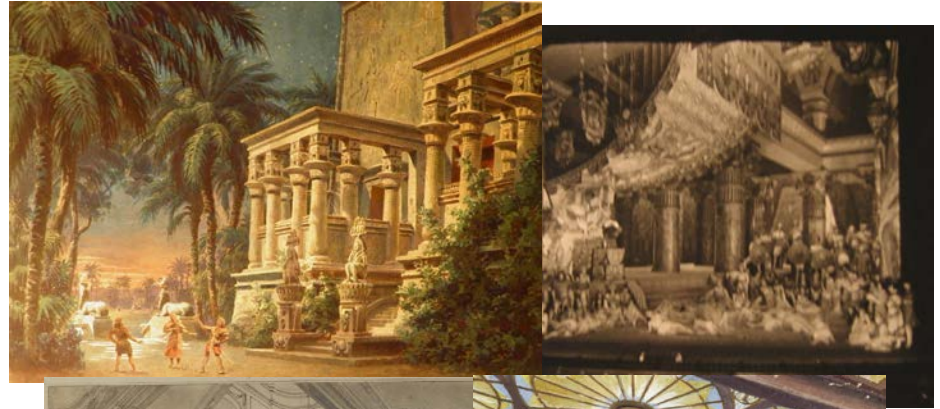
- Cercar diferències entre partitures
- Identificar l'escriptor (copista)



Fase 2.2: Extracció d'informació gràfica (2)

- Reconeixement de documents gràfics

- Escenografies
- Fotografies
- Il·lustracions
- Pintures
- Cartells



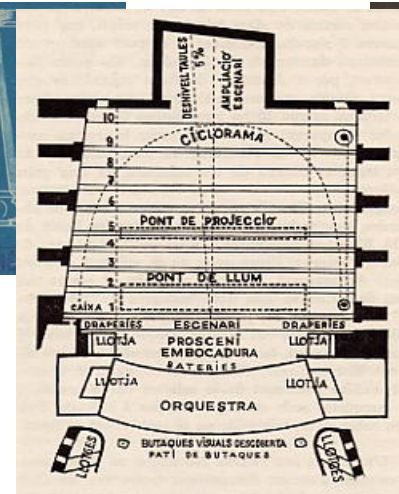
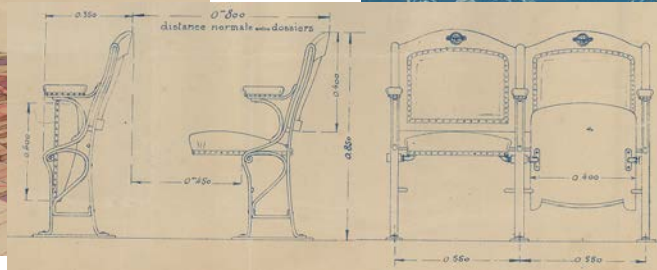
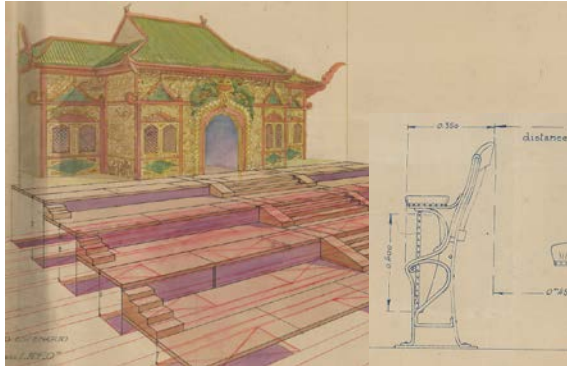
- Habilitació de cerques:

- Imatges similars
- Relacions entre imatges



Fase 2.2: Extracció d'informació gràfica (2)

- Plànols i Dibuixos



- Reconstrucció en 3 dimensions



Aplicacions

- Biblioteca Virtual Interactiva



- Facebook històric
 - Troba informació sobre una persona, esdeveniments i relacions



- Màquina del temps
 - Es troben i relacionen documents corresponents a una data concreta



Biblioteca Virtual Interactiva



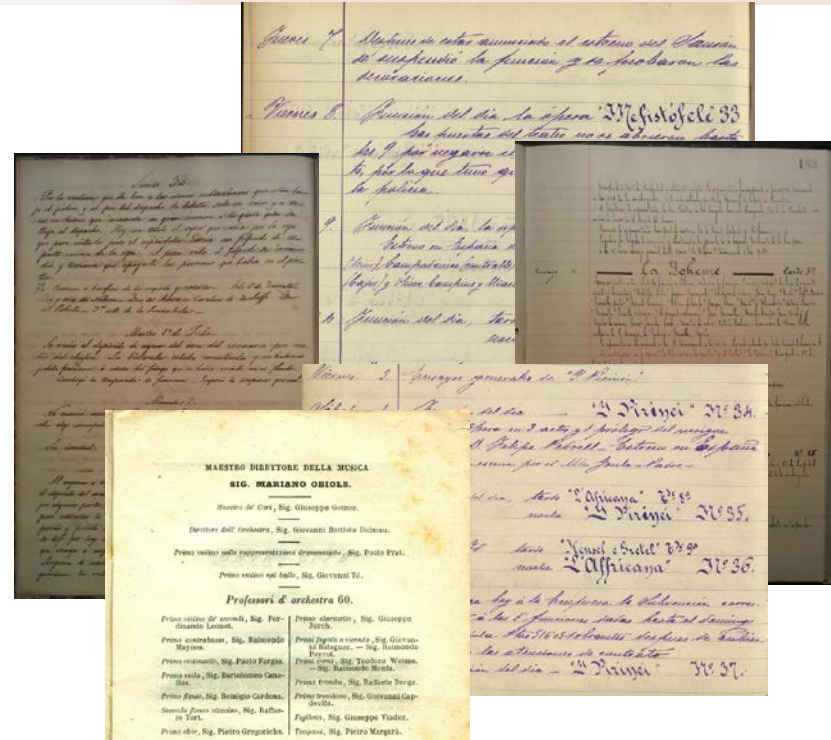
Facebook històric

Quina va ser la vida de Paolo Prat ?



- **Es fa la cerca entre els documents:**

- Diari del Conserge
- Actes de reunions
- Contractes / Pagaments
- Programes de mà, etc.



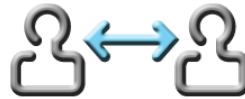
- **Reconstrucció de la seva vida:**



- Violinista, contractat l'any 1899.
- Va tocar a les següents òperes:

Tosca (gener 1899), *La Traviata* (març 1899), *Rigoletto* (juny 1900)...

- **Es troben relacions entre persones:**

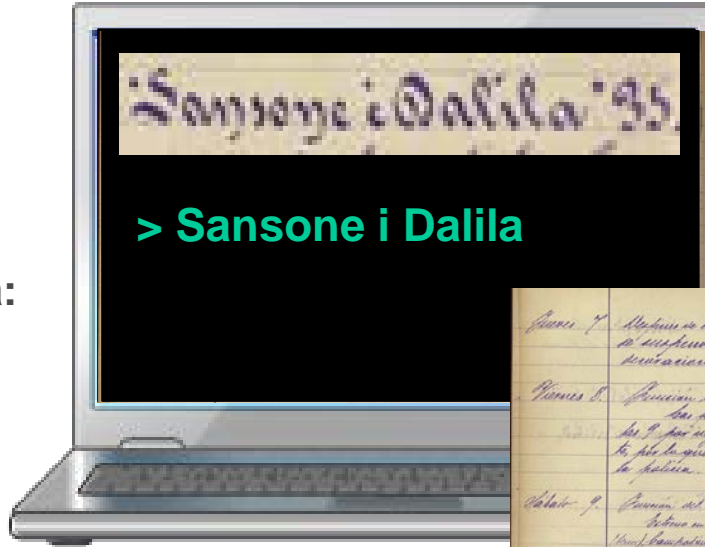


- Al gener de 1902 va coincidir amb el tenor **Luis Iribarne O'Connor** a l'òpera *Il Pirinei* de Felipe Pedrell.

Màquina del temps



9 de gener de 1897 ?

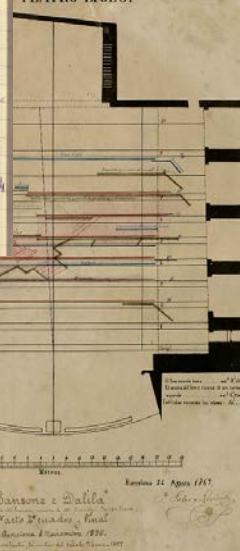
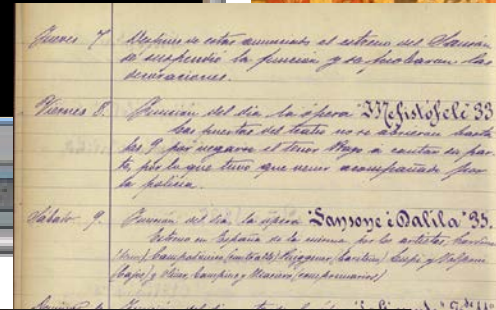


• Es fa la cerca automàtica:

- Diari del Conserge
- **Sanson i Dalila**
- Escenografies
- Cartells, Planols
- Fotografies, Il.lustracions
- Partitures Musicals

• Reconstrucció dels fets

- Video en 3D



Gràcies!