

# Content search with Pertimm

*Symposium on Digital Semantic  
Content across Cultures*

At C2RMF – Louvre – Paris

*May 5th 2006*

# Pertimm presentation

- Pertimm is a 1 M€ company specialized in generalized search engines
- Strong experience in linguistic, semantic, A.I., search engines, data bases, operating systems and content image retrieval
- Key actor in the Cap Digital French innovation Pole
- Many customers in France such as CNRS, ANPE, French Army...
- A big customer in U.S.: NASA

# Content search

- Linguistic search
- Semantic search
- Content image search

# Linguistic Search defined

- Defined as: working with words and their equivalent forms, I.E. using lexicon or dictionaries
- Example: augment, augmenting, augmentation, augmented, augmentations
- Wrong: cell = prison = cellular telephone = compartment

# Linguistic Search pros

- Experience shows an average of 4 times more pertinent that regular full text search

Mean number of forms	Probability of finding the occurrence	1 word request	2 words request	3 words request	4 words request	10 words request
2	English	50 %	25 %	13 %	3 %	-
4	Spanish/Italian	25 %	6 %	2 %	-	-
2	Japanese	50 %	25 %	13 %	3 %	-
3	French	33 %	11 %	4 %	-	-
6	German	17 %	3 %	-	-	-
8	Russian	13 %	2 %	-	-	-

# Linguistic Search example

## Request in German (agglutinant language) : Fahrzeug

Ländern zunehmend stabilisiert werden, steigen sie in den weniger entwickelten Ländern weiterhin an. Gemeinschaftsrichtlinien zur Festlegung strengerer Normen für die Schadstoffemissionen von **Kraftfahrzeugen** erbrachten positive Ergebnisse, aber die bislang erreichten Fortschritte drohen durch die zunehmende Zahl von **Fahrzeugen** auf den Straßen und durch die **Fahrzeugnutzung** zunichte gemacht zu werden. In den letzten Jahren hat der Kraftstoffverbrauch in der Gemeinschaft um jährlich 1,5% zugenommen.

Auf Gemeinschaftsebene sind mehrere Richtlinien erlassen worden, um die Verschmutzung durch den Verkehr zu verringern, sie legen Emissionshöchstgrenzen für **Fahrzeuge** und andere Quellen fest und sehen steuerliche Maßnahmen im Verkehrssektor vor, durch die der Verbraucher dazu gebracht werden soll, sich umweltfreundlicher zu verhalten.

### a. Auto-Öl-Programm

In Zusammenarbeit mit der Erdölindustrie und der Automobilindustrie hat die Kommission ein gemeinsames Programm, das sogenannte Auto-Öl-Programm, ausgearbeitet, durch das die Abgasemissionen verringert werden sollen. Dieses Programm besteht aus zwei neuen Richtlinien (98/69 und 98/70 zur Änderung der Richtlinien 70/156 und 70/220), die sich auf die Qualität von Benzin und Dieselmotorkraftstoff und auf Maßnahmen zur Bekämpfung der Luftverschmutzung durch Kraftfahrzeugemissionen beziehen.



- Durch die neue Richtlinie 98/70 soll die Verschmutzung durch Emissionen von **Kraftfahrzeugen** verringert werden, sie legt dafür neue Umweltspezifikationen für Benzin und Dieselmotorkraftstoff fest. Verbleites Benzin wird ab dem Jahr 2000 verboten. Vorgesehen sind außerdem schrittweise Verbesserungen der Umweltqualität von unverbleitem Benzin und von Dieselmotorkraftstoff.
- Die Richtlinie 98/69 legt unterschiedliche Emissionsgrenzwerte für Benzin- und **Dieselmotorkraftstoffe** fest, die je nach **Fahrzeugart** ab 2000 bzw. 2005 gelten sollen. Von den Mitgliedstaaten gewährte steuerliche Anreize, um eine frühere Einhaltung der neuen Grenzwerte zu fördern, sind zulässig, und ab dem 28. September 1999 wird der neue europäische Testzyklus, der in dieser Richtlinie vorgesehen ist, Anwendung finden.

# Semantic Search viewpoint

- Concept is defined as the object considered by a person plus the viewpoint of the person
- Thus, there are at least billions of concepts
- Each concept is in some way related to any other concept
- Thus, there billions of billions of billions, etc. of relations between concepts
- Thus semantic networks, or ontologies need to be virtualized as much as possible

# Semantic Search example

Evolutive Pertimm Interface

[» Reset interface](#)   [» Help](#)

**Pertimmizer search**

**Accuracy :** 67 (42 results)

1  100

**Search options**

**Answers to query :**

**investir capital risque fusion acquisition { on Europe }**

1-10 on 42 files found, sort by date | **pertinence** ▼ | barycenter | size | occurrences

« first   1 2 3 4 5 next »   end »

**1. EUROPEAN PARLIAMENT FACT SHEETS - 3.3.2. ABUSE OF A DOMINANT POSITION AND INVESTIGATION OF MERGERS**

Doc topics : concept of the dominant position, dominant positions, Steel Community Treaty, Community market, position of economic strength, Article 82 of the EC Treaty, market concerned, Basic Treaty provision, Article 82 of the Treaty, concept of abuse

[Highlighted document](#)

Date : 27/06/2002      Size : 23882      Pertinence : 100%      Format : TEXT      Language : English

**CONTEXT ASSOCIATIONS**

**TERM ANALYSIS**

investir (31, 33)  
capital (211, 643)  
risque (145, 306)  
fusion (95, 176)  
acquisition (138, 302)



# Semantic Search results

English:

□ (100) a. Company mergers, by concentration or acquisition, can obviously create or strengthen a dominant position likely to give rise to abuse. This risk justifies the Community authority in exercising prior control on merger operations. But while the ECSC Treaty grants the Commission exclusive power, under Article 66, to authorise or prohibit mergers between coal or steel companies, the EEC Treaty made no provision. The increase in mergers as a result of completion of the common market led to a need for Community intervention. ☑ At first this took the form of interpreting the existing provision

German:

□ (100) a. Ein Unternehmenszusammenschluss (durch Fusion oder Kauf) kann natürlich eine beherrschende Stellung begründen oder verstärken, die zu einer missbräuchlichen Ausnutzung führen kann. Angesichts dieses Risikos prüft die Gemeinschaftsbehörde Zusammenschlüsse im Voraus. Im Gegensatz zum EGKS-Vertrag (Artikel 66), der der Kommission die ausschließliche Befugnis einräumt, Zusammenschlüsse von Kohle- oder Stahlunternehmen zu gestatten oder zu verbieten, sieht der EWG-Vertrag nichts dergleichen vor. Durch die stärkere Annäherung der Unternehmen im Zuge des Binnenmarktes jedoch notwendig geworden. ☑ Zu diesem Zweck wurde

Spanish:

□ (100) a. Una concentración de empresas (por fusión o adquisición) evidentemente puede crear o reforzar una posición dominante que puede dar lugar a abusos. Este riesgo justifica la aplicación de un control previo de las operaciones de concentración por las autoridades comunitarias. Pero, a diferencia del Tratado CECA (artículo 66), que otorga a la Comisión el poder exclusivo de autorizar o prohibir las concentraciones de empresas, el Tratado CEE no había previsto nada al respecto. Ahora bien, el aumento de las fusiones de empresas acarreado por la realización del mercado comunitario. ☑ Ello se hizo en primer lugar a través de

Italian:

☑ (66) f. L'iniziativa "Innovazione 2000" (i2i) Lanciata nel marzo 2000, essa mira a costruire un'Europa fondata sulla conoscenza e l'innovazione. I prestiti hanno per oggetto i seguenti obiettivi: sviluppo delle PMI e dello spirito imprenditoriale, diffusione dell'innovazione, ricerca e sviluppo, reti di tecnologie dell'informazione e della comunicazione, arricchimento del capitale umano e investimenti in progetti audiovisivi europei. Dal maggio 2000 questi prestiti hanno riguardato le reti di comunicazione (965 milioni di euro), l'istruzione (448 milioni) e il capitale di rischio (214 milioni) a favore delle PMI innovatrici. ☑

# Content Image Search

- Pictorial technique selection

Texte recherché

Lancer la Recherche

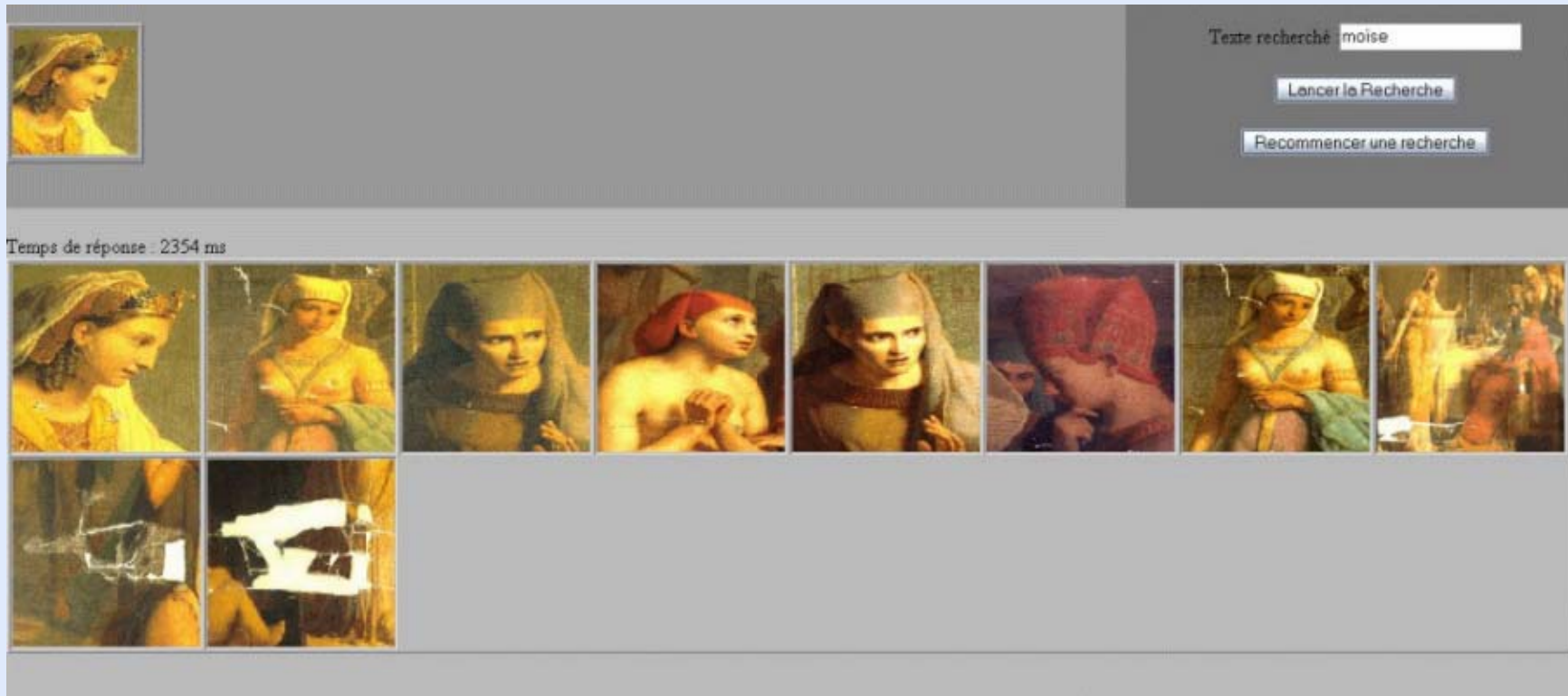
Recommencer une recherche

Temps de réponse : 1512 ms

The screenshot displays a search interface with a search bar containing the text 'Texte recherché'. Below the search bar are two buttons: 'Lancer la Recherche' and 'Recommencer une recherche'. To the left of the search bar, there are two small thumbnail images. Below the search bar, a response time of 'Temps de réponse : 1512 ms' is shown. The main area of the interface is a grid of 24 image thumbnails arranged in 3 rows and 8 columns. The images are diverse, including portraits, religious figures, and abstract compositions.

# Content Image Search

- Art history



Texte recherché : moise

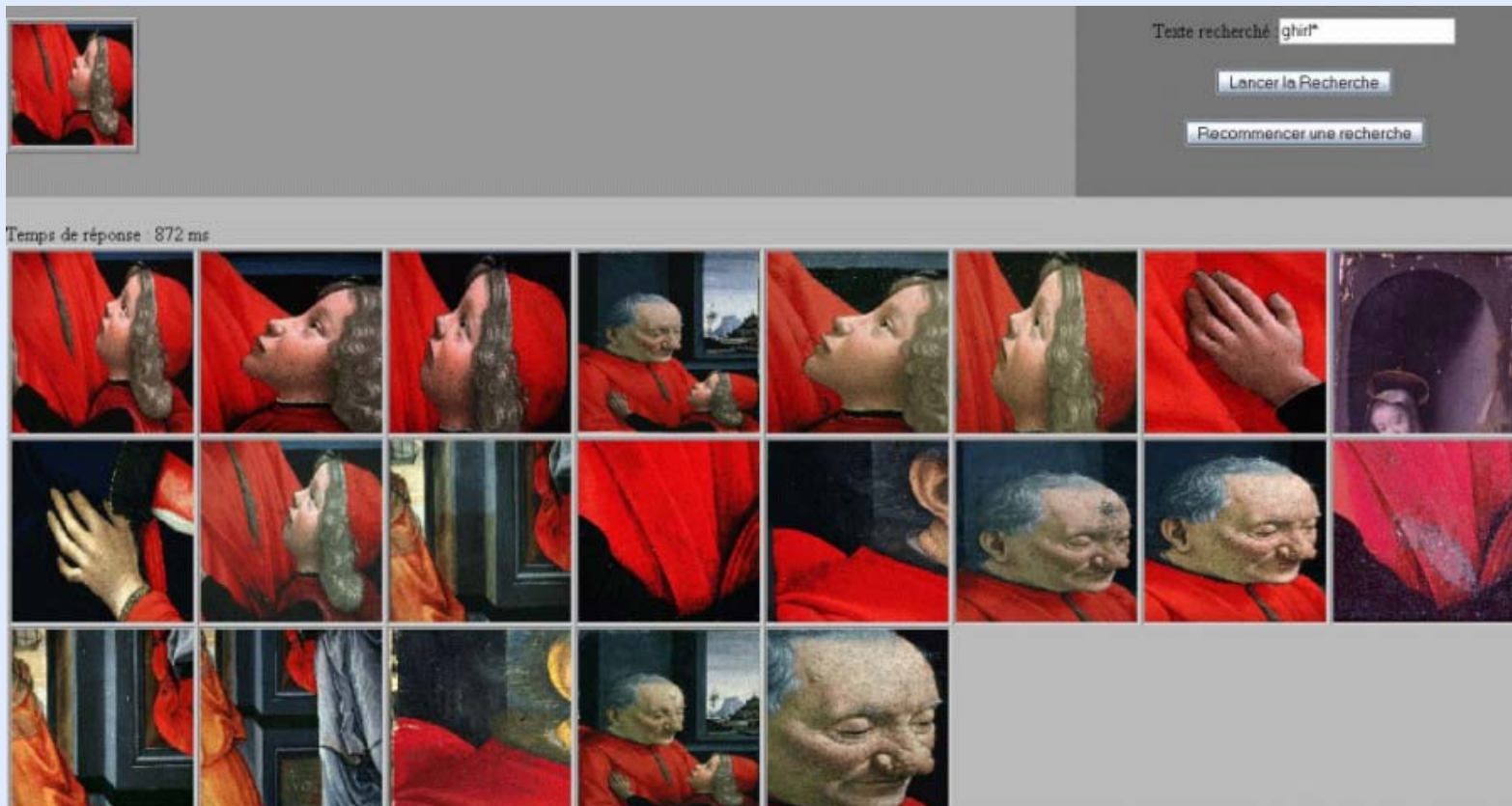
Lancer la Recherche

Recommencer une recherche

Temps de réponse : 2354 ms

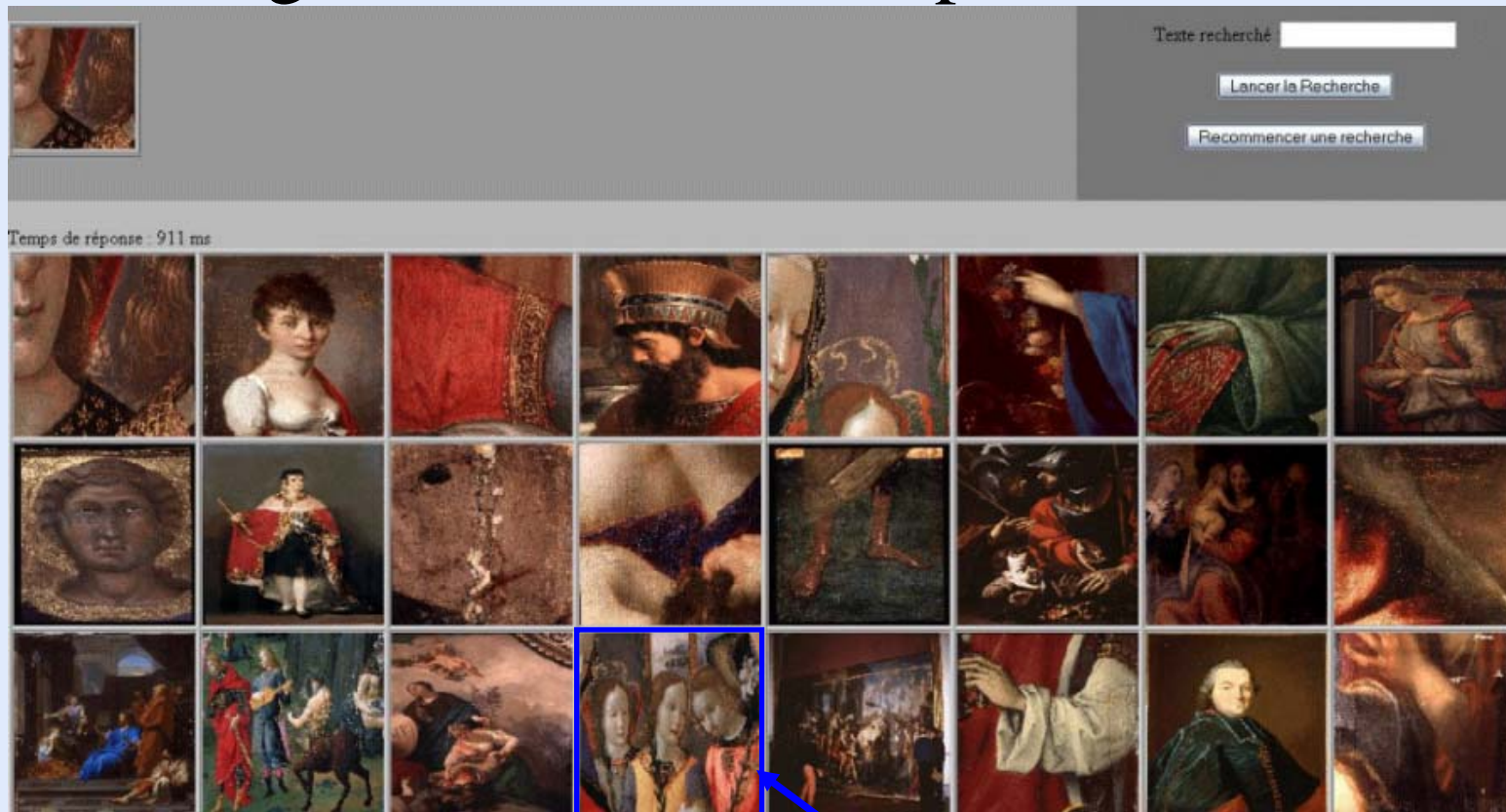
# Content Image Search

- Text at the service of the artwork



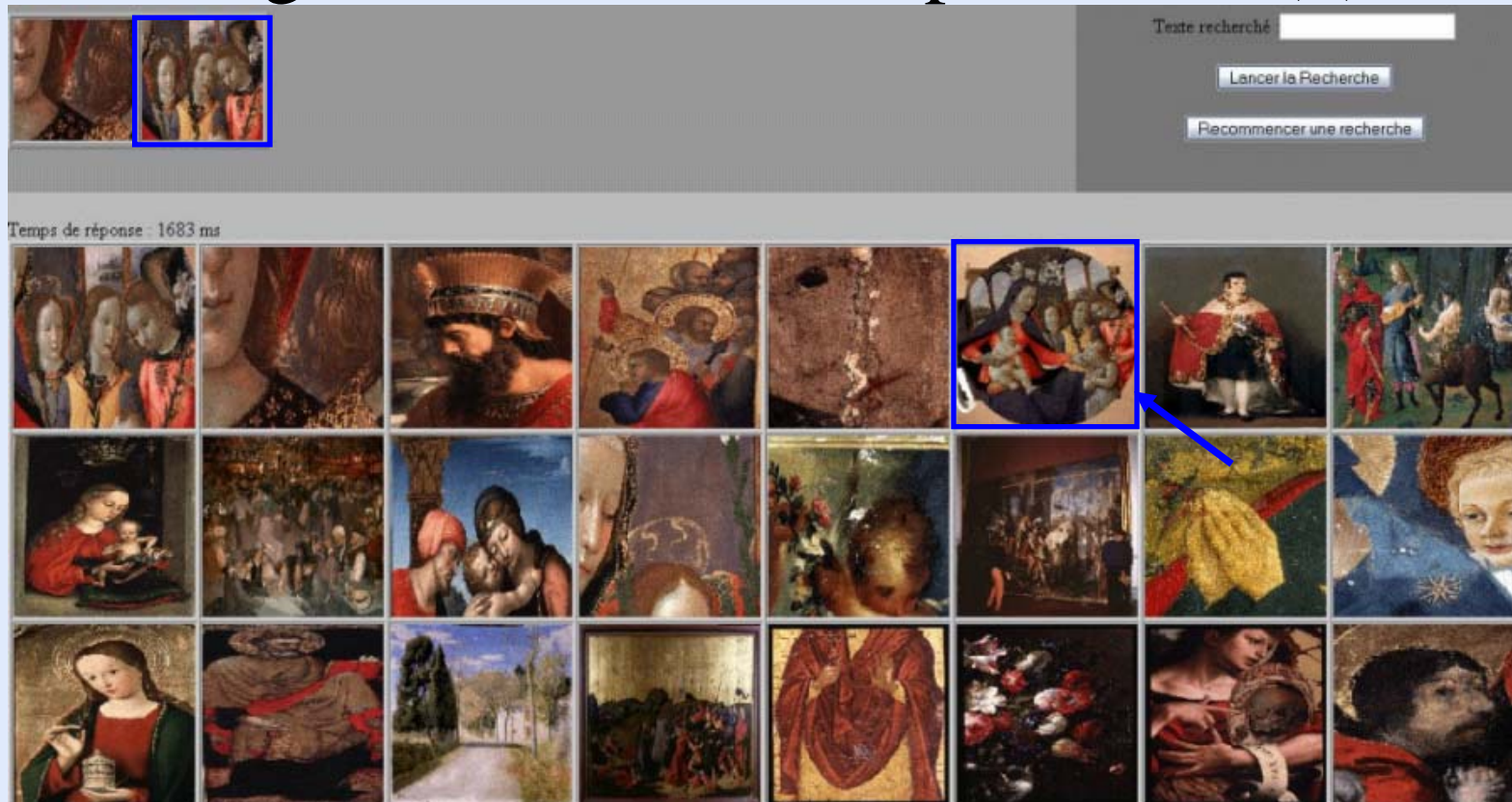
# Content Image Search

- Finding an artwork with a piece of it



# Content Image Search

- Finding an artwork with a piece of it (2)



# Epilogue

- Now: 100 000 images per machine, 10 millions documents per machine
- Project: 10 times better
- Unlimited is our motto