

# Repérage automatique d'informations temporelles touristiques dans des pages Web

Institut d'électronique et d'informatique Gaspard-Monge  
Séminaire de l'équipe Informatique Linguistique  
8 décembre 2008

Stéphanie Weiser

MoDyCo UMR 7114 – Université Paris Ouest Nanterre La Défense

---

# Plan général

- Cadre du travail
- Réalisation technique
- Le temps sur le Web
- Sémiotique des pages Web

# Plan – cadre du travail

- **Cadre du travail**

- Thèse
- Projets ANR
- Exemples d'expressions
- Exemples de pages Web
- Ontologie
- Repérage et annotation
- Expressions temporelles liées au tourisme

- Réalisation technique
- Le temps sur le Web
- Sémiotique des pages Web

---

# Thèse de doctorat

- Repérage et typage d'unités spatio-temporelles pour la construction d'une plate-forme d'annotation sémantique automatique de pages web
- Sous la direction de :
  - Philippe Laublet  
LaLIC, Université Paris Sorbonne
  - Jean-Luc Minel  
MoDyCo, Université Paris ouest Nanterre la Défense

# Projet Eiffel

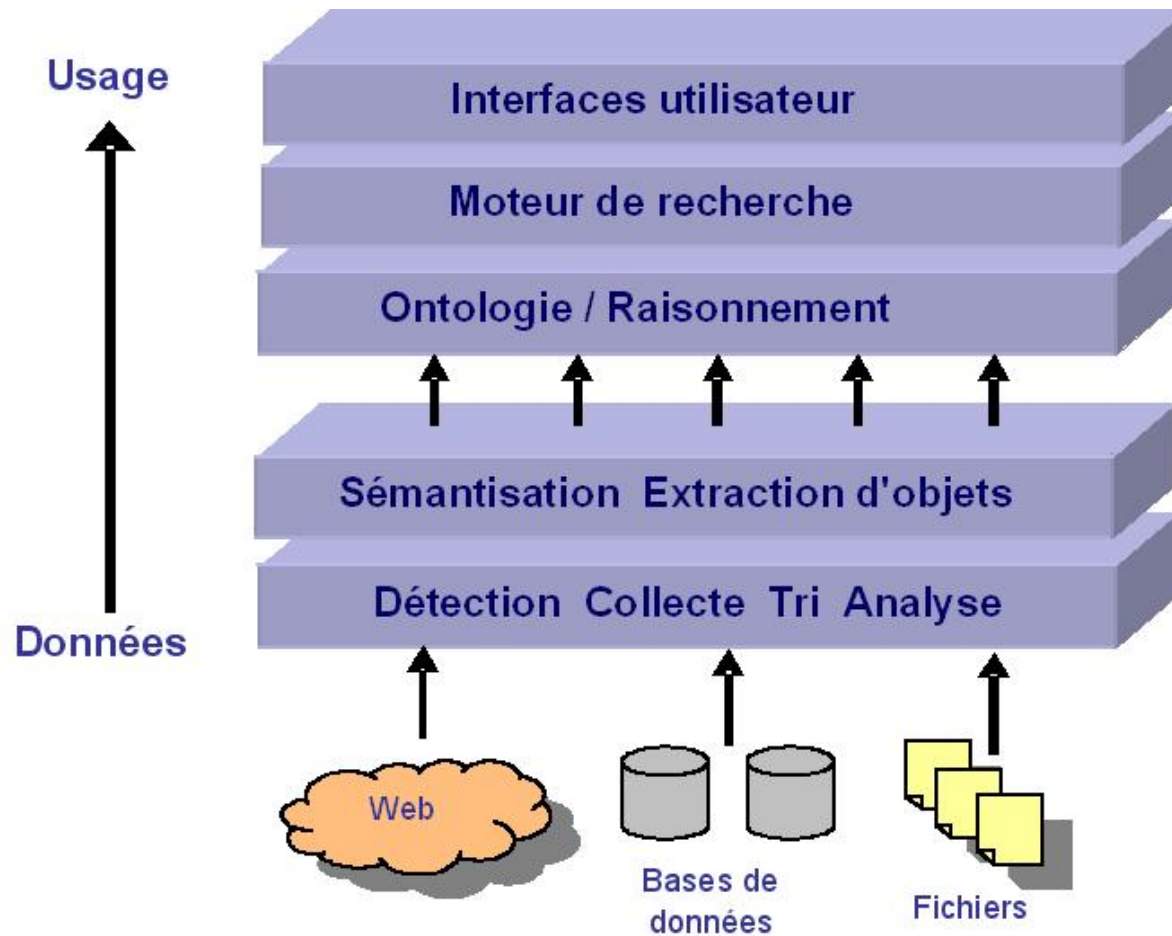
- eTourisme et Web Sémantique
- <http://www.projet-eiffel.org>
- ANR
- Antidot - Mondeca - Lirimm – Inria
- Donner à l'utilisateur une information riche et précise
- Permettre aux offreurs (territoires, opérateurs ou marchands) de mettre en avant leurs offres

---

# Projet TramedWeb

- Traces d'usage et médiations éditoriales dans les grands corpus du Web
- ANR Blanc
- Gripic, Culture et Communication
- Travaux plus théoriques, sémiotique

# Architecture d'Eiffel



---

# Exemples

- (1) Le 21 avril
- (2) Du 10 au 21 mars
- (3) Du lundi au vendredi, 9h – 11h
- (4) Inauguration du musée le 7 juillet 2006
- (5) En Juillet et Août ouvert tous les jours. Hors cette période, fermeture le mercredi soir, le jeudi toute la journée et le dimanche soir.  
Horaires d'ouverture : de 12h00 à 14h00 de 19h00 à 22h00.



# Exemples ambigus

- (1) Ouvert du lundi au vendredi de 9h à 11h et le dimanche en haute saison
- (2) Fermée les mardi, jeudi et samedi après-midi et dimanche
- (3) Horaires d'ouverture : lundi, mercredi 13h30\_17h30 mardi, jeudi, vendredi 8h-12h
- (4) Fermeture annuelle début Février à début Mars
- (5) Fermé une semaine en février

# Exemple

## LA MAIRIE



**MAIRIE D'OUROUX-EN-MORVAN**  
**58230 OUROUX-EN-MORVAN.**  
**Tél : 03 86 78 21 02 - Fax : 03.86.78.22 27**  
**E-mail : mairie.Ouroux-en-Morvan**

Fermée les mardi, jeudi et samedi après-midi  
et dimanche.

Fermer la fenêtre

# Exemple

## Auberge LA FERME DE CLAIREFONTAINE

Région : Centre



*Lettre d'information*

**Marie-Bernadette AVRAIN**

41160 SAINT-HILAIRE-LA-GRAVELLE  
Tél : 02 54 82 01 19 06 63 27 06 91 Fax : 02 54 82 06 91

Nous contacter » : [avrainsonia@neuf.fr](mailto:avrainsonia@neuf.fr)



### Ouverture :

Toute l'année, du jeudi au dimanche midi et soir sur réservation.  
Juillet et août tous les soirs et d'octobre à avril du vendredi au dimanche sauf groupes.

### Capacité :

**100 personnes**

**60 et 85 couverts pour groupe**

**Spécialités :** Foie gras parfum d'épices au miroir de fruits rouges, saumon mariné aux saveurs d'agrumes, délice de rouget barbet, sauce tropicale, salmis de canard, sabayon de poireaux, pigeonneau farce zéphir, sauté de veau aux pêches, civet de

# Exemple

## Office de Tourisme de Nevers et sa région\*\*\*



### Office de Tourisme de Nevers et sa région\*\*\* :

Palais ducal — BP 818 — Rue Sabatier

58008 NEVERS CEDEX

**Tél Accueil :** 03.86.68.46.00 — **Fax :** 03.86.68.45.98

#### ACCUEIL — INFORMATION

**Tél.** 03.86.68.46.00 - **fax.** 03.86.68.45.98

[contact@nevers-tourisme.com](mailto:contact@nevers-tourisme.com)

#### PROMOTION — RELATIONS PRESSE

**Tél.** 03.86.68.45.02 - **fax.** 03.86.68.45.03

[promo.presse@nevers-tourisme.com](mailto:promo.presse@nevers-tourisme.com)

#### RESA TOURISME

Organisation de visites pour groupes

**Tél.** 03.86.68.45.01 - 03.86.68.45.10 **fax.** 03.86.68.45.03

[resatourisme@nevers-tourisme.com](mailto:resatourisme@nevers-tourisme.com)

Une équipe dynamique à votre disposition toute l'année !

### HORAIRES 2007 :

#### Basse saison :

**Du 1er octobre 2006 au 31 mars 2007 :**

du lundi au samedi de 9h à 12h et de 14h à 18h

Fermé le dimanche et les jours fériés

#### Haute saison :

**Du 1er avril 2007 au 30 septembre 2007 :**

Du lundi au samedi de 9h à 18h30

Dimanche et jours fériés : 10h à 13h et de 15h à 18h

**L'Office de tourisme sera fermé les 1er janvier, 1er mai, 1er novembre, 11 novembre et 25 décembre 2007**

# Exemple

## **Piscine des Bords de Loire**

■ Rue Bernard Palissy 58000 NEVERS  
Tél. 03.86.71.83.99  
Piscine couverte.  
[s.sports@ville-nevers.fr](mailto:s.sports@ville-nevers.fr)

## **Piscine de la Jonction**

■ Quai de la Jonction 58000 NEVERS  
Tél. 03.86.37.55.95  
Piscine découverte. Piscine à vagues, Toboggan  
Ouverte en Juin, Juillet et Août

## **Le CRAPA Centre rustique d'activités physiques aménagés**

Parcours sportif de 1900 m ponctué de divers obstacles et exercices.  
■ Aux abords de l'étang de Niffonds à Varennes-Vauzelles.

## **CLUB VERT Centre de bien-être**

■ 21, rue Antony Duvivier à Nevers  
Tél. 03.86.21.56.62  
[www.clubvert-bienetre.com](http://www.clubvert-bienetre.com)  
E-mail : [info@clubvert-bienetre.com](mailto:info@clubvert-bienetre.com)  
Cardio training, musculation, fitness, aéro-biking, danse et sports de combat self  
défense, body vive.  
30 cours collectifs par semaine par professeurs diplômés d'état  
Espace détente sauna-hammam...

## **GOLF PUBLIC DU NIVERNAIS**

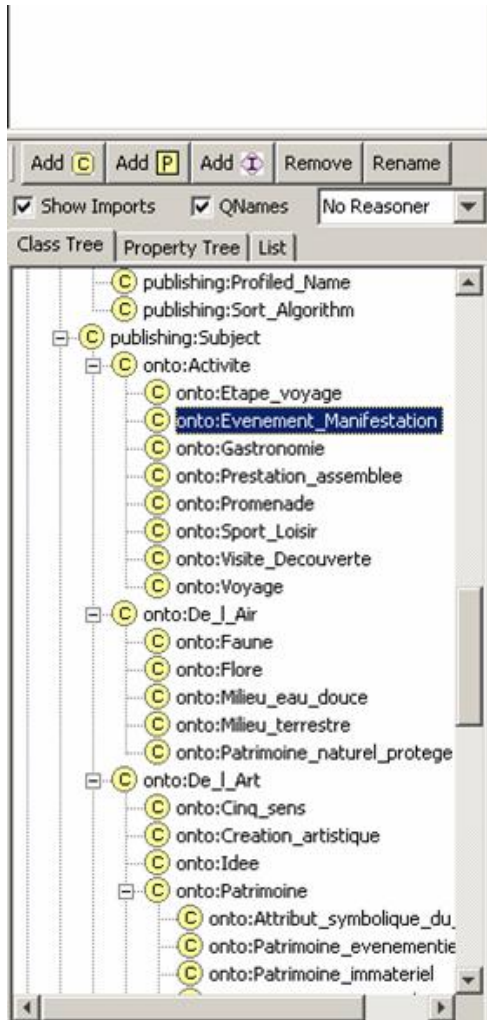
■ Le Bardonnay  
58470 MAGNY-COURS  
Tél. : 03.86.58.18.30 / Fax : 03.86.58.04.04  
[www.cg58.fr](http://www.cg58.fr) - [golf-public-du-nivernais@wanadoo.fr](mailto:golf-public-du-nivernais@wanadoo.fr)  
Le golf du Nivernais, situé en bordure du technopôle de Nevers -Magny-Cours, fait partie de l'espace -  
loisirs réalisé à proximité du circuit automobile de Formule1.  
Le golf du Nivernais offre ainsi à tous les passionnés d'automobile un parc de verdure de 50 ha et à  
tous les passionnés de golf un moment d'émotion intense sur la piste internationale de karting ou sur

# Ontologie (1)

- Point central :
  - ❑ Guide l'extraction
  - ❑ Guide l'annotation
  - ❑ Guide la base de connaissances
- Ontologie du tourisme et du territoire
  - ❑ qualification fine des ressources touristiques (activité, hébergement, logistique, patrimoine, territoires, voyages et itinéraires...)
  - ❑ Mise au point et adaptée selon les besoins des utilisateurs



# Ontologie (2)



## Annotations:

rdfs:comment (Datatype  
http://www.w3.org/2001/XMLSchema#string) : Fête / Manifestation  
artistique / Manifestation commerciale / Manifestation  
profes-sionnelle / Manifestation religieuse / Manifestation sportive /  
Manifestation traditionnelle /...  
rdfs:label (fr) : Manifestation  
basicontology:SortName (Datatype  
http://www.w3.org/2001/XMLSchema#string) : 61  
rdfs:comment (fr) : Événement ou manifestation

## Subclass of:

[onto:Activite](#)

## Domain of:

[onto:site\\_web](#)  
[onto:commune\\_at](#)  
[onto:description](#)  
[onto:adresse\\_postale](#)  
[onto:zone\\_geographique\\_at](#)  
[onto:periodiciteEvenement](#)  
[onto:e-mail](#)  
[onto:accueil](#)  
[onto:contact](#)  
[onto:label\\_at](#)  
[onto:telephone](#)  
[onto:visuel](#)  
[onto:bassin\\_touristique\\_at](#)  
[onto:date\\_fin](#)  
[onto:date\\_debut](#)  
[onto:organisateur](#)

# Repérage & annotation

- Repérage des informations temporelles, des informations spatiales et des objets touristiques dans les pages Web
- Annotation selon une DTD calquée sur l'ontologie et la base de connaissances
- Utilisation d'Unitex



# Caractérisation des expressions temporelles

- Repérage à l'aide de marqueurs lexicaux
- Informations pratiques et informatives
  - Horaires d'ouverture / de fermeture
    - *Ouvert de 9h à 18h*
  - Périodes d'ouverture / de fermeture
    - *Fermé du 12 au 16 janvier*
  - Exceptions
    - *Ouvert tous les lundis, sauf en été*
- Expressions absolues
  - *Le 12 janvier 2008*

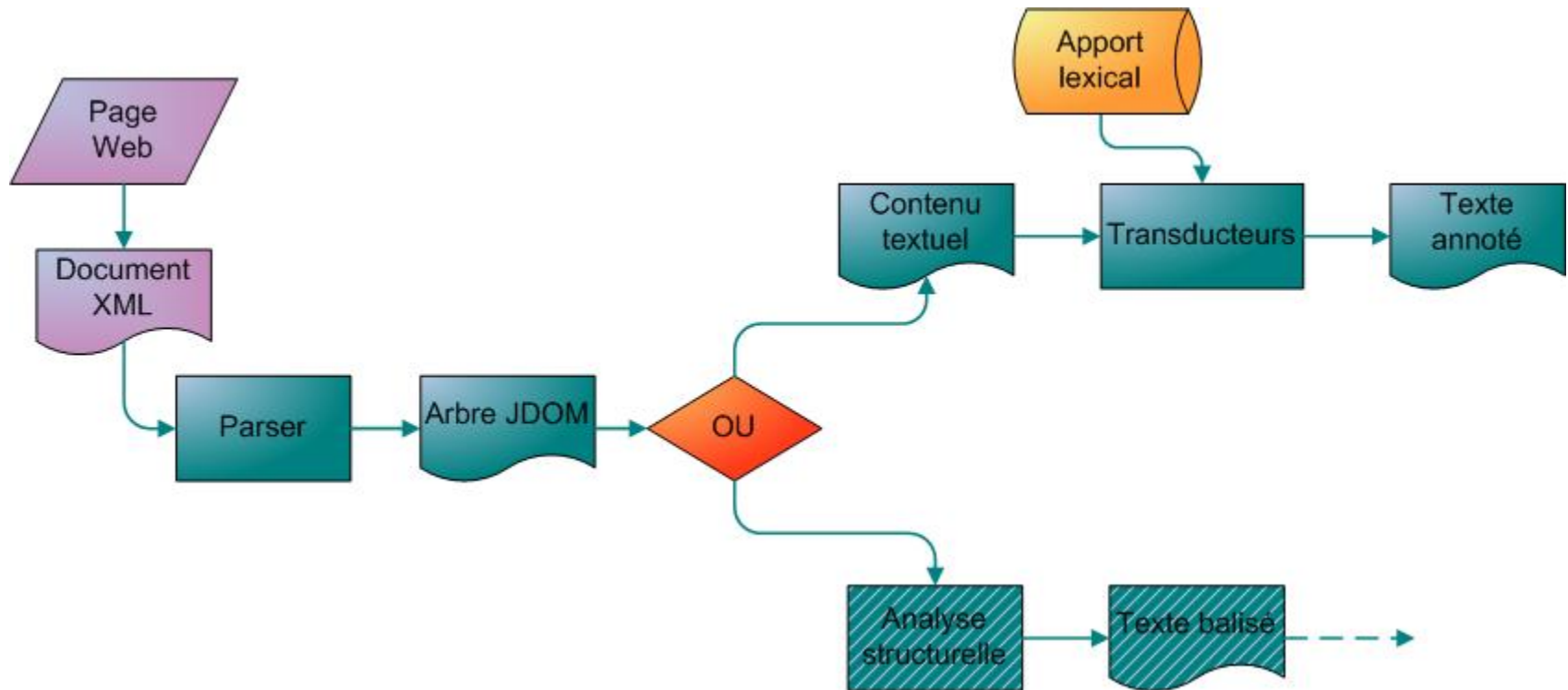
# Plan – réalisation technique

- Cadre du travail
- **Réalisation technique**
  - Architecture générale
  - Architecture du module d'annotation
  - Analyse linguistique
  - Transducteurs
  - Annotations
  - Inférences
- Le temps sur le Web
- Sémiotique des pages Web

# Architecture générale

- Antidot fournit les pages Web :
  - Collecte automatique de données touristiques non structurées sur le Web
  - Transformation du html en xml en :
    - Supprimant certaines balises
    - Gardant les marques de mise en forme
- MoDyCo repère et annote les informations temporelles
  - Fournit un fichier xml annoté

# Extraction – Annotation : Architecture



# Analyse linguistique

- Analyse manuelle de pages Web (100 pages)
- Comment sont formulées les expressions temporelles ?
- Étude de la disposition
- Marques linguistiques
- Élaboration en parallèle des transducteurs

# Transducteurs Unitex

- 24 transducteurs
  - 88 marqueurs de surface (*ouvert, mardi...*)
  - 22 marqueurs généraux (*le, de...*)
- Pas d'utilisation de graphes existants :
  - Trop généraux
  - But : cibler exactement nos expressions
- Pas d'utilisation de Time-MI ou autre :
  - Annotation d'événements dans des textes linéaires

# Annotations (dtd)

- Periode\_Ouverture / Periode\_Fermeture
  - Heure\_Debut / Heure\_Fin
  - Date\_Debut / Date\_Fin
  - Date
  - Jours
  - Partie\_Journee
- Incertitude
- Exception

# Annotations (choix)

- Granularité
  - Jour
  - Partie de journée (matin, après-midi, soir)
- Pas d'inférences
  - Annotation des ouvertures et des fermetures
  - Conversion pour la base de connaissance
- « Exception » : description textuelle
- Qualification de l'expression
  - Certaine ou incertaine



# Inférences

- Transformer les informations de fermeture en ouverture
  - Système à base de règle
  - Fermé le mardi
- Méthode probabiliste
  - Degré de possibilité d'ouverture ou de fermeture
  - Modélisation de connaissances vagues, ambiguës, contextuelles
  - Fermé de fin janvier à fin février

---

# Plan – le temps sur le Web

- Cadre du travail
- Réalisation technique
- **Le temps sur le Web**
  - Web participatif
  - Exemple Webcity
  - Comparaison guide papier – guide Web
- Sémiotique des pages Web

# Le temps sur le Web

- Impossible de généraliser « le Web »
  - Pages Web touristiques
    - Présentant un objet
    - Présentant plusieurs objets
    - Dates absolues
  - Exemples du Web participatif
    - Dates relatives

---

# Web participatif

- Exemple de Webcity
  - Mise à jour
  - Importance de la date de parution de l'information :
    - Logique de dépôt
    - Logique de consultation
    - Fiabilité

# Exemple Webcity

[Présentation](#) [Actu](#) [Mot du patron](#) [Photos](#) [Pratique](#) [Avis\\*](#)

### Détail des avis

Accueil/Service : ★★★★★ Décor/Ambiance : ★★★★★ Qualité/Prix : ★★★★★

Ce qu'ils en pensent :

Les hommes : ★★★★★ Les femmes : ★★★★★

Les - de 25 ans : ★★★★★ De 25 à 45 ans : ★★★★★ Les + de 45 ans : ★★★★★

Vous pouvez trier par : [note](#) ▼ , [Date](#) ▼



 **Audere** 

03/01/2008 ★★★★★

**Tranquille pour bien déjeuner**

Testé uniquement pour déjeuner : pas les uns sur les autres, service froid mais à

---

# Comparaison Guide Papier – Pages Web

- Modifiable
- Dates relatives
- Pas de régularité :  
ni dans la structure,  
ni dans les  
formulations
- Année de parution
- Stabilité de  
l'information
- Dates absolues ou  
pour l'année de  
parution
- Régularité

# Guide papier

50 LE PARIS MÉDIÉVAL : RIVE DROITE

## Où manger ?

### Sur le pouce

#### À la française

**Boulangerie Hervé Malineau** (plan, 4) : 26, rue Saint-Paul, 75004. ☎ 01-48-87-64-10. M Saint-Paul. Ouvert de 7 h à 20 h. Fermé le mercredi. Sandwichs à la painpaulette à 3,50 €. Si vous n'êtes pas encore un adepte de la painpaulette, c'est le moment de craquer

### Sur un plateau

#### À la française

**Le Dénicheur** (plan, 11) : 4, rue Tiquetonne, 75002. ☎ 01-42-21-31-01. M Étienne-Marcel. Ouvert tous les jours de 11 h à 15 h et de 19 h à minuit. Formules à 8,50 € le midi, 12 et 15,50 € le soir. À la carte, compter autour de 20 € pour un repas complet. Petit troquet rigolo, à la déco de bric et de broc, très

#### À l'américaine

**Scoop** (plan, 12) : 154, rue Saint-Honoré, 75001. ☎ 01-42-60-31-84. M Louvre-Rivoli ou Palais-Royal-Musée-du-Louvre. Ouvert tous les jours de 11 h à 18 h. Fermé le soir, sauf le vendredi. Formules entre 10 et 15 €. Brunch à 20 ou 23 € le dimanche de 11 h à 16 h. Un restaurant tonique, relaxant, en phase avec le monde et ses produits les plus naturels. Ce petit resto

pour cette nouvelle Reine du Marais qui mène à la baguette nombre de grands chefs parisiens. Découvrez-la dans les sandwichs poulet-crudités ou crudités-gruyère. Sinon, les sandwichs aux pavés Saint-Paul sont à 4,20 €.

*seventies* (ça fait aussi brocante). Petite bouffe simple et généreuse, autour de copieuses salades, d'assiettes de pâtes ou du plat du jour. Idéal pour les petites bourses ou pour se retrouver entre copains. Un café offert à nos lecteurs sur présentation de ce guide.

d'étage ne désemplit pas, depuis l'ouverture. La formule plaît autant pour sa carte que pour ses prix. Ou pour son cadre : large baie vitrée donnant sur le temple de l'Oratoire, briques jaunes ou pierres apparentes, tables basses et banquettes, coin-snack coloré. Des soupes originales, des cakes salés, des plats du jour qui font voyager sur le pouce. Glaces originales. Un *Scoop*

qui tient ses promesses. Coin-épicerie au rez-de-chaussée !

**Foody's Salad Bar** (plan, 13) : 26, rue Montorgueil, 75001. ☎ 01-40-13-02-53. M Les Halles. Ouvert du lundi au samedi de 11 h à 17 h 30. Trois menus, ou plutôt des formules, de 8 à 12,50 €. Voici une idée très californienne dans l'esprit. Il s'agit de proposer des produits très frais et

#### À l'orientale

**La Galerie 88** (plan, 15) : 88, quai de l'Hôtel-de-Ville, 75004. ☎ 01-42-72-17-58. M Pont-Marie ou Hôtel-de-Ville. Ouvert tous les jours de 10 h à 2 h ; service continu jusqu'à 1 h. Menu à 10 €, servi midi et soir. Compter 15 € à la carte. Entre Marais et île Saint-Louis, une

#### En terrasse

**Lézard Café** (plan, 20) : 32, rue Étienne-Marcel, 75002. ☎ 01-42-33-22-73. M Étienne-Marcel. Ouvert de 9 h à 2 h. Fermé le dimanche. Salades autour de 10 € et plats de 10 à 13 €. Un lieu qui devrait servir un jour de décor pour

#### Avec un p'tit verre

**À la Cloche des Halles** (plan, 25) : 28, rue Coquillière, 75001. ☎ 01-42-36-93-89. M Louvre-Rivoli ou Les Halles. Ouvert de 8 h à 22 h (de 10 h à 16 h le samedi). Fermé le dimanche et les 3 premières semaines d'août. Compter entre 6,50 et 8 € pour une assiette de

légers, cuisinés simplement mais avec précision. On se sert donc (libre-service), en soupe, en salades variées et inventives, en dessert et en boisson avant d'aller à l'une des petites tables en bois. Le midi, celles et ceux qui ont peu de temps pour déjeuner mais veulent le faire de manière saine et sans s'alourdir trouveront là leur bonheur.

petite étape relax pour déguster un thé au milieu de souvenirs et d'objets rapportés de lointains pays. Ça donne un côté chaleureux bien plaisant. Petite restauration, salades, *mezze* et pâtes, mais ne rêvez pas de grandes découvertes culinaires.

raconter les Halles de notre temps. Finis les clichés à la Doisneau. La déco comme la clientèle font désormais dans la sobriété. Ici, on déjeune ou on dîne tard. Du bon, du frais, qu'ils disaient. Y'a pas d'lézard... sur la terrasse.

charcuterie ou de fromage. Verre de vin de 1,45 à 4,40 €. Une cloche qui, du temps des Halles, annonçait le début et la fin des marchés, surplombe la façade. Au son de cloche, on pouvait récupérer ce qui restait. Pas la peine de vous ruer pour autant sur le comptoir, les jours

BALADE N° 3 51

BALADE N° 3

# Observations

- « Cadre » : rubrique restaurant
- Régularité : pour chaque entrée :
  - Nom
  - Adresse
  - Horaires d'ouverture ...
- Structuration !



# Page Web correspondante



# Page Web correspondante



## Scoop [sku:p]

n. masc. (mot angl.) : nouvelle importante publiée en exclusivité  
n. fém. (mot angl.) : grande cuillère



011254

154, rue Saint Honoré 75001 Paris - Tél. : 01 42 60 31 84  
Métros : Louvre/Rivoli ou Palais Royal

scoopcafe®  
Tous droits réservés

---

# Observations

- Pas de « cadre » :
  - Comment savoir qu'il s'agit d'un restaurant ?
- Pas de régularité entre les pages
  - Pas la même structuration
  - Pas les mêmes informations
- Images et disposition font sens

---

# Plan – Sémiotique des pages Web

- Cadre du travail
- Réalisation technique
- Le temps sur le Web
- **Sémiotique des pages Web**
  - Étude de la disposition
  - Tests de déplacement
  - Observations

---

# Sémiotique des pages Web

- Étude de la disposition
- Éléments déplaçables
- Pas de linéarité
- Test

### *Quatre chambres d'hôtes dans un manoir*

<a href="#">Accueil</a>	<a href="#">Capacité, tarifs</a>	<b>Bienvenue</b>  chez Paul et Bernadette Colas Chaumotte 58120 St Hilaire en Morvan  Tel : 33 (0)3 86 85 22 33
<a href="#">Le manoir</a>	<a href="#">Réservation</a>	
<a href="#">Chambres</a>	<a href="#">Plan du coin</a>	
<a href="#">Restauration</a>	<a href="#">Sites à proximité</a>	
<a href="#">Loisirs</a>	<a href="#">Actualités</a>	
<a href="#">Sites touristiques</a>	<a href="#">Photos</a>	

### *Bienvenue dans une ferme d'élevage de moutons du Parc Naturel Régional du Morvan*



Nous parlons anglais  
*We speak english*



**Ouvert du 15 Février au 30 Novembre**

Pour nous envoyer un mail : [cliquez ici](#)

### *Quatre chambres d'hôtes dans un manoir*

<a href="#">Accueil</a>	<a href="#">Capacité, tarifs</a>
<a href="#">Le manoir</a>	<a href="#">Réservation</a>
<a href="#">Chambres</a>	<a href="#">Plan du coin</a>
<a href="#">Restauration</a>	<a href="#">Offres de proximité</a>
<a href="#">Loisirs</a>	<a href="#">Actualités</a>
<a href="#">Sites touristiques</a>	<a href="#">Photos</a>

### **Bienvenue**

chez Paul et Bernadette Colas  
Chaumotte  
58120 St Hilaire en Morvan  
Tel : 33 (0)3 86 85 22 33

### *Bienvenue dans une ferme d'élevage de moutons du Parc Naturel Régional du Morvan*



Nous parlons anglais  
*We speak english*



**Ouvert du 15 Février au 30 Novembre**

Pour nous envoyer un mail : [cliquez ici](#)



*Bienvenue dans une ferme d'élevage de moutons  
du Parc Naturel Régional du Morvan*



*Quatre chambres d'hôtes dans un manoir*

**Ouvert du 15 Février au 30 Novembre**



Nous parlons anglais  
*We speak english*



## Bienvenue

chez Paul et Bernadette Colas  
Chaumotte  
58120 St Hilaire en Morvan  
Tel : 33 (0)3 86 85 22 33

Accueil

Le manoir

Chambres

Restauration

Loisirs

Sites touristiques

Capacité, tarifs

Réservation

Plan du coin

Gîtes à proximité

Actualités

Photos

Pour nous envoyer un mail : [cliquez ici](#)



Ouvert du 15 Février au 30 Novembre

*Bienvenue dans une ferme d'élevage de moutons  
du Parc Naturel Régional du Morvan*



<u>Accueil</u>	Capacité, tarifs	<b>Bienvenue</b>  chez Paul et Bernadette Colas Chaumotte 58120 St Hilaire en Morvan  Tel : 33 (0)3 86 85 22 33
Le manoir	Réservation	
Chambres	Plan d'accès	
Restauration	Gîtes à proximité	
Loisirs	Actualités	
Sites touristiques	Photos	

*Quatre chambres d'hôtes dans un manoir*



Nous parlons anglais



*We speak english*

Pour nous envoyer un mail : [cliquez ici](#)

# Observations

- Sens conservé
- Choix des blocs significatifs
- Pas de linéarité
- Image (auberge)

# Conclusion

- Expressions propre au Web
  - Floues et ambiguës
- Pas seulement des expressions textuelles
- Prise en compte de la disposition
- Projet :
  - Système permettant d'utiliser la disposition

---

# Questions ?