

Université de Montréal

Présentation personnalisée des informations environnementales

par
Mohamed Mouine

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en informatique

Juin, 2014

© Mohamed Mouine, 2014.

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Présentation personnalisée des informations environnementales

présentée par:

Mohamed Mouine

a été évaluée par un jury composé des personnes suivantes:

Jian-Yun Nie,	président-rapporteur
Guy Lapalme,	directeur de recherche
El Mostapha Aboulhamid,	membre du jury
Laurence Danlos,	examineur externe

Thèse acceptée le:

RÉSUMÉ

Nous présentons dans cette thèse notre travail dans le domaine de la visualisation. Nous nous sommes intéressés au problème de la génération des bulletins météorologiques. Étant donné une masse énorme d'information générée par Environnement Canada et un utilisateur, il faut lui générer une visualisation personnalisée qui répond à ses besoins et à ses préférences.

Nous avons développé MeteoVis, un générateur de bulletin météorologique. Comme nous avons peu d'information sur le profil de l'utilisateur, nous nous sommes basés sur les utilisateurs similaires pour lui calculer ses besoins et ses préférences.

Nous utilisons l'apprentissage non supervisé pour regrouper les utilisateurs similaires. Nous calculons le taux de similarité des profils utilisateurs dans le même cluster pour pondérer les besoins et les préférences.

Nous avons mené, avec l'aide d'utilisateurs n'ayant aucun rapport avec le projet, des expériences d'évaluation et de comparaison de notre outil par rapport à celui utilisé actuellement par Environnement Canada. Les résultats de cette évaluation montrent que les visualisation générées par MeteoVis sont de loin meilleures que les bulletins actuels préparés par EC.

Mots clés: visualisation de l'information, personnalisation, profil utilisateur, apprentissage machine, extraction de donnée, graphique.

ABSTRACT

We present our work in this thesis in the field of information visualization. We dealt with the problem of the generation of weather forecasts reports. Given the huge amount of information produced by Environment Canada and a wide variety of users, it must generate a customized visualization that meets their needs and preferences.

We developed MeteoVis, a weather report generator. Given that we have little information on the user profile, we relied on the choices made by similar users to calculate the needs and preferences of a user.

We use unsupervised machine learning techniques to group similar users . We compute a degree of similarity of user profiles in the same cluster to determine the needs and preferences.

We conducted, with the help of external users experiments for evaluating and comparing our tool with the current site of Environment Canada. The evaluation results show that the visualizations generated by MeteoVis are significantly better than the current bulletins prepared by EC.

Keywords : information visualization, personalization, user profile, machine learning, data mining, graph.

TABLE DES MATIÈRES

RÉSUMÉ	iii
ABSTRACT	iv
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX	viii
LISTE DES FIGURES	ix
CHAPITRE 1 : PRÉSENTATION DU PROBLÈME	1
1.1 Flux d'information à Environnement Canada	1
1.1.1 Stations météorologiques	2
1.1.2 SCRIBE	4
1.2 Description du fichier MeteoCode	6
1.3 Fichiers Citypage	7
1.4 Comparaison MeteoCode et Citypage	10
1.5 Personnalisation de la visualisation	11
1.6 Conclusion	13
CHAPITRE 2 : ÉTAT DE L'ART	14
2.1 Génération des graphes	15
2.2 Techniques de visualisation	20
2.3 Perception	20
2.3.1 Traitement préattentif	22
2.3.2 Les principes de Gestalt	23
2.4 Interactivité	25
2.5 Génération de texte	28
2.6 Génération combinée des textes et des graphiques	29
2.7 Analyse de l'existant	31
2.8 La théorie confrontée à la pratique	33
2.8.1 Environnement Canada	35
2.8.2 NOAA	38
2.8.3 MétéoFrance	40
2.8.4 Meteoblue	43
2.8.5 Moteur de recherche	45

2.8.6	Synopsis	45
2.9	Conclusion	47
CHAPITRE 3 : PERSONNALISATION PAR CLUSTERING		49
3.1	Introduction	49
3.2	Profil de l'utilisateur	49
3.3	Notre proposition	52
3.3.1	Critères et paramètres	54
3.3.2	Clustering	56
3.3.3	Calcul de similarité	59
3.3.4	Calcul des préférences	59
3.4	Génération de la visualisation	60
3.5	Evaluation et amélioration de la visualisation	62
3.5.1	Analyse de sensibilité du seuil	63
3.6	Conclusion	64
CHAPITRE 4 : METEOVIS		66
4.1	Architecture	66
4.2	Extraction des données	67
4.3	Génération de la visualisation	75
4.3.1	Clustering	78
4.3.2	Condition courante	78
4.3.3	Graphique	79
4.3.4	Génération de texte	83
4.4	Conclusion	86
CHAPITRE 5 : ÉVALUATION ET AMÉLIORATION		87
5.1	Interactivité	88
5.2	Principes de Gestalt	91
5.3	Étude empirique : Rétroaction des utilisateurs	93
5.3.1	Scénarios	95
5.3.2	Évaluation du rendement de l'utilisateur	96
5.4	Évaluation de l'expérience utilisateur	99
5.5	Interprétation et amélioration	100
5.5.1	Seuil de décision	103
5.6	Conclusion	104

CHAPITRE 6 :	CONCLUSION ET TRAVAUX FUTURS	105
6.1	Travaux futurs	107
6.1.1	Évaluation publique	107
6.1.2	Génération de texte	108
6.1.3	Application à d'autres domaines	110
6.1.4	Médecine	110
6.1.5	Finance	111
6.2	Conclusion	113
BIBLIOGRAPHIE		114

LISTE DES TABLEAUX

1.I	Comparaison entre MeteoCode et Citypage	10
2.I	Tableau récapitulatif	48
3.I	Critères du profil utilisateur	54
3.II	Taux de similarité	60
3.III	Résultat du calcul des préférences	61
3.IV	Résultat de l'interaction avec la visualisation	62
3.V	Résultat final de la variation des seuils (après 3 itérations)	64
4.I	Avantages et inconvénients des deux méthodes d'extraction de données proposées	70
4.II	Les intervalles pour la couverture nuageuse	83
4.III	Exemple d'apprentissage des gabarits de phrases utilisés par EC	85
5.I	Tableau récapitulatif des techniques d'interactivité incluant Me- teoVis	89
5.II	Tableau récapitulatif des principes de Gestalt incluant MeteoVis	91
5.III	Résultat de l'évaluation	102

LISTE DES FIGURES

1.1	Flux d'information simplifié à Environnement Canada	3
1.2	Exemple de fichier MeteoCode	5
1.3	Exemple de fichier Citypage	8
1.4	Le site officiel d'Environnement Canada : http://www.meteo.gc.ca	9
1.5	Exemple de visualisation générée	12
2.1	Les 5 graphiques les plus utilisés	16
2.2	Les fronts météorologiques	19
2.3	Les symboles des directions du vent	19
2.4	Les symboles météo	19
2.5	Un exemple de recherche d'un cercle cible rouge basée sur une différence de couleur	23
2.6	Architecture d'un système de génération de langue naturel (Reiter et Dale, 2000)	30
2.7	Les étapes de l'approche de Agrawala et al. (2011) pour la conception d'une visualisation	32
2.8	Visualisation des informations météorologiques (RALI)	34
2.9	Le site officiel d'Environnement Canada : http://www.meteo.gc.ca	37
2.10	NOAA : visualisation et zoom des avertissements	39
2.11	NOAA : Prévisions détaillées	40
2.12	Page d'accueil de Météo France	41
2.13	Prévisions des précipitations de 1 heure	42
2.14	Visualisation générée par Meteo Blue utilisant les pictogrammes et le rainSPOT	44
2.15	Visualisation générée par Meteoblue utilisant des symboles météorologiques	44
2.16	Résultats affichés par différents moteurs de recherche	46
3.1	Notre approche pour créer des visualisations personnalisées	52
3.2	Exemple de visualisation générée	55
4.1	Architecture du système MeteoVis	68
4.2	Schéma de la base de données	71
4.3	Lien d'un fichier MeteoCode	72
4.4	Extrait de la liste des codes des fichiers MeteoCode	73
4.5	Aperçu de MeteoVis dans une téléphone mobile	77

4.6	La condition courante générée par MeteoVis	79
4.7	Le graphique généré par MeteoVis	79
4.8	Le vent	81
4.9	La température	82
4.10	Présentation de l'accumulation et du type de la précipitation . . .	82
4.11	Le texte généré par MeteoVis	84
5.1	Exemple d'infobulle.	91
5.2	Modèle d'évaluation et de raffinement	96
5.3	Questionnaire pour l'évaluation de la visualisation	98
6.1	Exemple de code Java de SimpleNLG-ENFR permettant de générer la phrase « On prévoit des vents forts d'ouest à 90 km/h ou plus accompagnés de pluie. » Nous donnons en paramètres les valeurs de la vitesse et la direction du vent et le type de précipitation. On prépare la description du vent auquel on ajoute la précipitation pour réaliser la phrase. Nous remercions Pierre Luc Vaudry pour son aide dans la réalisation de cet exemple.	109
6.2	Calendrier économique	112

CHAPITRE 1

PRÉSENTATION DU PROBLÈME

Ces dernières années, il y a eu une explosion du volume de données générées dans tous les domaines de la connaissance, la tâche la plus difficile étant devenue leur analyse et leur exploration. La fouille de données nous permet de localiser les informations nécessaires. La visualisation de l'information et l'exploration visuelle des données peuvent aider à mieux assimiler l'information quand elle est combinée avec une description textuelle. Le processus de visualisation d'information et de données scientifiques cherche à résoudre le problème de représentation des différents types de données pour l'utilisateur, de sorte que les données peuvent être facilement communiquées et interprétées.

Le RALI¹ est impliqué dans un projet² en collaboration avec Environnement Canada, Notre partenaire dans ce projet produit une masse énorme d'informations météorologiques de façon continue. Dans la section suivante nous présentons certaines composantes du système d'information au sein d'EC en insistant sur les parties reliées à notre projet.

1.1 Flux d'information à Environnement Canada

La durée de vie d'une information météorologique est très courte. L'information brute générée par les stations météo passe par plusieurs étapes dans le système d'EC avant de devenir « consommable ». Dans cette section, nous expliquons le cycle de vie d'une information météorologique chez EC. La figure 1.1 montre, de gauche à droite, une vue simplifiée de la circulation de l'information chez EC, allant de la prévision

1. Le RALI réunit des informaticiens et des linguistes d'expérience dans le traitement automatique de la langue. Il est le plus important laboratoire universitaire dans le domaine au Canada.

2. <http://rali.iro.umontreal.ca/EnvironmentalInfo/index.fr.html>

numérique de la météo aux bulletins météorologiques en français et en anglais publiés régulièrement, ainsi que les avertissements qui sont produits lors de circonstances météorologiques exceptionnelles ou quand des événements spéciaux de l'environnement se produisent, par exemple, les tempêtes, la grêle, la pluie verglaçante, etc. Ces entrées et sorties sont indiquées en gras dans la figure 1.1.

1.1.1 Stations météorologiques

Notre projet cherche à s'intégrer au système d'information d'Environnement Canada illustré à la figure 1.1. Les sources d'informations sont les stations météorologiques qui fournissent des mesures physiques intégrées à des modèles mathématiques et des simulations numériques pour produire des prédictions de valeurs de paramètres météorologiques. Parmi les paramètres enregistrés on retrouve la température, la vitesse et la direction du vent, la couverture nuageuse, le type (neige ou pluie) et la valeur de la précipitation et l'accumulation qu'on peut considérer exactes. Les valeurs de prédiction résultent d'un calcul complexe selon plusieurs scénarios. Toutes les stations météorologiques génèrent et envoient leurs rapports au moins une fois par heure³ à Environnement Canada.

À partir des informations de ces stations, un modèle de prévision numérique produit une grande quantité de données toutes les 12 heures : 6 Mb pour chacune des 5 régions du Canada et 9 Mb d'information globale pour l'ensemble du Canada. Ces données sont stockées dans différents formats, l'un d'entre eux étant des matrices Scribe contenant les paramètres météorologiques estimés comme la vitesse des vents, pressions, températures, précipitations et d'autres dans 800 points de toutes les régions du

3. Environnement Canada a conclu des accords avec l'Organisation météorologique mondiale qui permettent d'effectuer des observations une fois l'heure en période de beau temps. Lorsque la météo devient plus active et qu'on prévoit des vents violents, de la grêle ou de la pluie verglaçante, la fréquence des observations augmente afin de mieux alerter le public des changements climatiques imprévus et potentiellement dangereux (source : <http://past.meteomedia.com/index.php?product=help&pagecontent=forecasts#Q14>)

1.1.2 SCRIBE

SCRIBE (Verret et al. (1997)) est un système interactif qui a été développé par le centre météorologique canadien pour la composition de produits de prévisions météorologiques à partir de matrices des éléments météorologiques disponibles à un ensemble de stations ou points d'échantillonnage.

Les trois principales composantes fonctionnelles de ce système à base de connaissances sont le compilateur de règles, le moteur d'inférence pour interroger les règles et le système de gestion de la base de faits. Ces deux derniers éléments sont combinés pour extraire et manipuler les concepts intégrés dans les données brutes, basés sur la connaissance représentée par un ensemble de règles. Environ 600 règles constituent la base de connaissances utilisée pour générer les concepts, et plus de 1200 règles constituent la base de connaissances pour générer divers produits finaux de prévisions.

Lors de la réception des matrices, SCRIBE traite les données pour extraire les événements météorologiques ou les concepts résultant d'une analyse numérique sémantique du contenu des matrices d'éléments météorologiques.

Une autre sortie de Scribe est un ensemble de fichiers XML appelé Meteocode. Ce format ouvert et validable avec un schéma XML donne des informations détaillées sur les événements météorologiques dans chaque région ou groupe de points géographiques. Ces données sont également mises à la disposition des clients d'EC⁴ qui souhaitent développer leurs propres applications ou prévisions.

Jusqu'à cette étape tout le travail est réalisé automatiquement sans (ou presque) intervention humaine. L'intervention des météorologues permet de générer :

- MeteoCode, généré sous forme XML (voir figure 1.2) et aussi sous forme de fichier plat (csv), contient les prévisions pour les sept prochains jours. Les informations sont détaillées heure par heure. Ce fichier ne contient pas de description

4. <http://dd.weatheroffice.gc.ca/meteocode/>

Figure 1.2 – Exemple de fichier MétéoCode. Les informations qui apparaissent dans cet exemple sont : <head> contient la meta-information du fichier (catégorie, date de création, période de validité...), <location> la zone géographique concernée par ces informations et <parameters> les paramètres de la météo comme la température, la précipitation, la couverture nuageuse...

```

    <more-information>http://weatheroffice.ec.gc.ca</more-information>
  <production-center>...</production-center>
  <feedback/>
  <credit/>
  <credit-logo/>
  <disclaimer>...</disclaimer>
</source>
</head>
<data>
  <forecast>
    <meteocode-forecast>
      <location>
        <msc-zone-code>r71.1</msc-zone-code>
        <msc-zone-name lang="fr">Montréal métropolitain - Laval</msc-zone-name>
        <msc-zone-name lang="en">Metro Montréal - Laval</msc-zone-name>
      </location>
      <parameters>
        <warning-list/>
        <cloud-list units="deci">...</cloud-list>
        <precipitation-list/>
        <probability-of-precipitation-list units="%">...</probability-of-precipitation-list>
        <accum-list units="mm">...</accum-list>
        <snow-level-list units="m"/>
        <temperature-list type="air" units="celsius">...</temperature-list>
        <temperature-list type="dew-point" units="celsius">...</temperature-list>
        <temperature-list type="climatology" units="celsius">...</temperature-list>
        <wind-list units="kmh">...</wind-list>
        <visibility-list/>
        <UV-index-list>...</UV-index-list>
        <temperature-list type="sea-surface" units="celsius"/>
        <freezing-spray-list/>
        <wave-height-list units="m"/>
        <ice-cover-list units="%"/>
      </parameters>
    </meteocode-forecast>
  </forecast>
  <meteocode-forecast>
    <location>
      <msc-zone-code>r71.2</msc-zone-code>
      <msc-zone-name lang="fr">Vaudreuil - Soulanges - Huntingdon</msc-zone-name>
      <msc-zone-name lang="en">Vaudreuil - Soulanges - Huntingdon</msc-zone-name>
    </location>
  </meteocode-forecast>
</data>

```

textuelle de l'état climatique, ce qui explique la génération d'une seule version.

Nous utilisons ces fichiers pour générer notre visualisation.

– CityPage⁵ Un fichier XML (voir figure 1.3) contenant les conditions courantes

5. http://dd.weatheroffice.gc.ca/citypage_weather/xml/

de la météo et les prévisions pour les sept prochains jours. Les informations de ce fichier indiquent les valeurs des paramètres pour chaque jour (exemple : Lundi température min= -6, température max= 3). Une description textuelle est générée accompagnant les valeurs numériques. Ce qui explique la génération de deux versions de ce fichier (français et anglais) qui sont utilisées pour préparer ses bulletins climatiques.

- Les avertissements : Si les conditions météorologiques risquent de présenter un risque pour la sécurité des personnes et/ou des biens, les météorologues d'Environnement Canada rédigent des avertissements visant à prévenir les personnes se trouvant dans des zones concernées.

1.2 Description du fichier MeteoCode

Le fichier MeteoCode (figure 1.2) contient des données détaillées. Les prévisions des 7 prochains jours sont réparties sur deux fichiers MeteoCode. Le premier appelé meteo-code02 (contenant les détails des conditions actuelles et des prévisions des deux jours suivants) et meteocode37 (contenant les prévisions des quatre jours d'après). Chaque fichier MeteoCode contient les informations météorologiques d'une région. Ces informations sont spécifiques à plusieurs emplacements dans cette région. Le contenu d'un élément `<meteocode-forecast>` contient les informations météorologiques à un emplacement précis (voir figure 1.2). Nous pouvons donc extraire les informations pertinentes en recherchant dans l'arbre de `<meteocode-forecast>` identifié par le code correspondant de la région de la ville `<msc-zone-code>`. Pour chaque emplacement, nous pouvons extraire les informations suivantes :

`<cloud-list>` un code de 0 à 10 (10 : totalement nuageux) est attribué à chaque heure de la journée.

`<precipitation-list>` : type de précipitation, l'heure de début et l'heure de la fin.

<probability-of-precipitation> : l'heure de début et l'heure de la fin et la probabilité (en pourcentage) d'avoir des précipitations.

<accum-list> : Accumulation totale, accumulation horaire et type (pluie ou neige) de l'accumulation pour chaque heure (vide s'il n'y a pas d'accumulation).

<temperature-list> : la température de l'air, point de rosée et climatologique pour chaque heure.

<wind-list> : Vitesse, direction, heure de début et heure de fin du vent.

<visibility-list> : degré de visibilité si un brouillard est prévu.

<UV-index-list> : indice des rayons ultra violet du soleil.

1.3 Fichiers Citypage

Les fichiers Citypage (figure 1.3) sont préparés pour les bulletins météorologiques qu'on trouve sur le site d'EC (voir figure 1.4).

EC publie deux fois par jour 837 fichiers CityPage dans chaque langue (français et anglais). Ce nombre peut paraître énorme et donner l'impression que ce fichier est très détaillé avec une prévision précise dans le temps et dans l'espace. Or, compte tenu de la taille du Canada, ces bulletins sont agrégés et ne présentent pas tous les détails disponibles dans le MeteoCode. Ces fichiers sont le résultat d'une agrégation dans l'espace et dans le temps des informations contenues dans les fichiers MeteoCode. Les seules informations contenues dans les bulletins d'EC sont :

- La condition courante : température, point de rosée, pression, visibilité, humidité et la vitesse et la direction du vent.
- Prévision des sept prochains jours : température minimum et température maximum pour chaque jour et une icône indiquant la couverture nuageuse prévue pour sept jours.

Figure 1.3 – Exemple de fichier CityPage : les information qui apparaissent dans cette exemples sont : les details de la date <dateTime> , les details de l'emplacement <location> , les valeurs de la codition courante de la météo <currentConditions> et les prévisions des jours suivants <forecastGroup> .

```
<siteData xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation=
  <license>http://dd.weather.gc.ca/doc/LICENCE_GENERAL.txt</license>
  <dateTime name="xmlCreation" zone="UTC" UTCOffset="0">
    <year>2013</year>
    <month name="September">09</month>
    <day name="Monday">02</day>
    <hour>23</hour>
    <minute>20</minute>
    <timeStamp>20130902232000</timeStamp>
    <textSummary>Monday September 02, 2013 at 23:20 UTC</textSummary>
  </dateTime>
  <dateTime name="xmlCreation" zone="EDT" UTCOffset="-4">...</dateTime>
  <location>
    <continent>North America</continent>
    <country code="ca">Canada</country>
    <province code="qc">Quebec</province>
    <name code="s0000635" lat="45.52N" lon="73.65W">Montréal</name>
    <region>Metro Montréal - Laval</region>
  </location>
  <warnings url="http://weather.gc.ca/warnings/report_e.html?qc67">
    <event type="warning" priority="urgent" description="SEVERE THUNDERSTORM WARNING IN EFFECT">
      <dateTime name="eventIssue" zone="UTC" UTCOffset="0">...</dateTime>
      <dateTime name="eventIssue" zone="EDT" UTCOffset="-4">...</dateTime>
    </event>
    <event type="watch" priority="high" description="SEVERE THUNDERSTORM WATCH IN EFFECT">...</
  </warnings>
  <currentConditions>
    <station code="yul" lat="45.47N" lon="73.74W">Montréal-Trudeau Int'l Airport</station>
    <dateTime name="observation" zone="UTC" UTCOffset="0">...</dateTime>
    <dateTime name="observation" zone="EDT" UTCOffset="-4">...</dateTime>
    <condition>Thunderstorm with light rainshowers</condition>
    <iconCode format="gif">19</iconCode>
    <temperature unitType="metric" units="C">19.9</temperature>
    <dewpoint unitType="metric" units="C">18.9</dewpoint>
    <pressure unitType="metric" units="kPa" change="0.19" tendency="falling">100.6</pressure>
    <visibility unitType="metric" units="km">12.9</visibility>
    <relativeHumidity units="%">94</relativeHumidity>
    <wind>...</wind>
  </currentConditions>
  <forecastGroup>
    <dateTime name="forecastIssue" zone="UTC" UTCOffset="0">...</dateTime>
    <dateTime name="forecastIssue" zone="EDT" UTCOffset="-4">...</dateTime>
    <regionalNormals>
      <textSummary>Low 11. High 22.</textSummary>
      <temperature unitType="metric" units="C" class="high">22</temperature>
```

- Description textuelle : une brève description textuelle accompagne les deux précédentes parties décrivant l'état de la météo en une dizaine de mots.

L'utilisateur venant sur le site d'Environnement Canada doit passer par quelques

Figure 1.4 – Le site officiel d'Environnement Canada : <http://www.meteo.gc.ca>. Pour arriver à cette page (une ville précise), l'utilisateur choisit la langue (français ou anglais) puis la ville sur une carte ou dans une liste.

The screenshot shows the Environment Canada website interface. At the top, there's a navigation bar with 'Météo', 'Alertes', 'Maritime', 'Qualité de l'air', 'Données', 'Archives', and 'Recherche'. A large blue arrow points from the 'Météo' section to a map of Canada. Another blue arrow points from the map to the 'Montréal, QC' link. Below the map, there's a section for 'Conditions actuelles' (Current Conditions) for 'Aéroport int. de Montréal-Trudeau' on '2 septembre 2013'. The current temperature is 20°C. Below this, there's a 'Prévisions' (Forecast) section showing a 7-day forecast from Sunday to Sunday. The forecast includes icons for weather conditions, high and low temperatures, and a percentage chance of precipitation. Below the forecast, there's a section for 'Émissions à: 11h30 HAE le lundi 2 septembre 2013' (Emissions at: 11:30 AM EDT on Monday, September 2, 2013). This section provides detailed weather information for 'Aujourd'hui' (Today), 'Ce soir et cette nuit' (Tonight and tonight), and the following days (Mardi, Mercredi, Jeudi, Vendredi, Samedi, Dimanche).

Conditions actuelles 24 dernières h | Radar | Satellite

Enregistrées à: **Aéroport int. de Montréal-Trudeau**
 Date: **12h00 HAE le lundi 2 septembre 2013**

Condition : **Nuageux** Température : **20,0°C**
 Pression : **100,8 kPa** Point de rosée : **18,0°C**
 Tendence : **à la baisse** Humidité : **88%**
 Visibilité : **19 km** Vent : **NNE 12 km/h**
 Humidex : **26**

Prévisions CAS

lun	mar	mer	jeu	ven	sam	dim
2 sep	3 sep	4 sep	5 sep	6 sep	7 sep	8 sep
23°C	21°C	23°C	17°C	23°C	22°C	20°C
18°C	14°C	8°C	10°C	16°C	12°C	

Émissions à: 11h30 HAE le lundi 2 septembre 2013

Aujourd'hui
 Nuageux. Averses ou orages débutant tard cet après-midi. Hauteur prévue de 30 mm sous les orages. Maximum 23. Indice UV de 4 ou modéré.

Ce soir et cette nuit
 Averses ou orages cessant vers minuit. Partiellement nuageux par la suite. Minimum 18.

Mardi
 Généralement nuageux. 40 pour cent de probabilité d'averses en après-midi et en soirée. Vents devenant du sud-ouest à 20 km/h avec rafales à 40 le matin. Maximum 21.

Mercredi
 Alternance de soleil et de nuages. Minimum 14. Maximum 23.

Jeudi
 Ensoleillé. Minimum 8. Maximum 17.

Vendredi
 Nuageux. Minimum 10. Maximum 23.

Samedi
 Averses. Minimum 16. Maximum 22.

Dimanche
 Alternance de soleil et de nuages. Minimum 12. Maximum 20.

étapes (p.e. choisir la province et la ville) avant d’arriver au bulletin qu’il désire.

Les bulletins préparés actuellement par EC présentent certaines limitations que nous voulons traiter :

- Ces textes sont limités à quelques dizaines de mots trouvés dans les prévisions météorologiques régionales.
- Les icônes utilisées sont génériques pour toute la journée.
- L’information présentée est la même pour tous les usagers.

On peut remarquer que ces limitations sont dues essentiellement à deux raisons : EC prépare ces bulletins au préalable et ne prend pas en considération que l’emplacement par l’utilisateur. Beaucoup d’information est perdue par l’agrégation dans le temps et dans l’espace pour créer les fichiers CityPage, source d’information de ces bulletins.

1.4 Comparaison MeteoCode et Citypage

Contrairement à Citypage , l’information contenue dans MeteoCode est complète et contient tous les détails disponibles. Le tableau 1.I compare les deux types de fichiers.

Tableau 1.I – Comparaison entre MeteoCode et Citypage

Paramètre	MeteoCode	Citypage
Conditions courantes	oui	oui
Prévision 7j	oui	oui
Température	air, point de rosée, climatologique	air
Température à chaque heure	oui	-
Couverture nuageuse	oui	oui
Couverture nuageuse h/h	oui	-
Probabilité de precipitation	indique l’heure de début et celle de la fin	par jour
Vent	l’heure de début et celle de la fin	par jour
Description textuelle	-	oui

L’information diffusée sur le site d’EC est celle contenue dans le fichier Citypage. L’utilisateur ne bénéficie pas de la granularité disponible dans MeteoCode. EC est conscient

qu'une grande partie de l'information sur les prévisions n'est pas diffusée de façon adéquate sur son site. C'est dans ce cadre qu'a été établie une entente de contribution pour étudier différentes méthodes pour améliorer la présentation des informations environnementales. En particulier notre projet qui vise à développer des méthodes et des approches pour générer des rapports visuels pour cette grande masse d'information évoluant dans le temps et l'espace tout en tenant compte des particularités et de la localisation de l'utilisateur.

1.5 Personnalisation de la visualisation

Étant donné l'étendue du territoire canadien, EC ne peut préparer au préalable des bulletins spécifiques pour chaque besoin. La solution étudiée dans cette thèse est de générer des bulletins météorologiques sur demande incluant seulement les informations dont l'utilisateur a besoin. Cette approche permet à l'usager de récupérer facilement l'information qui l'intéresse sans avoir à consulter toute la masse d'information d'EC.

Le Météocode est une donnée ouverte au format XML validé, donc est facilement analysable. Nous nous sommes concentré sur la présentation la plus appropriée des données. Étant donnée la taille des données, nous développons des techniques particulières pour l'agrégation des données dans l'espace et dans le temps.

Comme nous traitons des données XML, nous avons d'abord étudié l'utilisation des feuilles de style XML (XSLT) pour sélectionner et produire des pages Web, mais étant donné la taille des données, la simple sélection et l'affichage des données ne sont pas suffisantes.

À partir des informations du MeteoCode, nous avons développé MeteoVis⁶ un générateur de bulletins climatiques pour une adresse ou un code postal donné par l'utilisateur. MeteoVis produit des bulletins sur demande. Le contenu du bulletin résultant est

6. <http://www-etud.iro.umontreal.ca/~mouinemo/meteo/meteovis.php>

une combinaison de texte et de graphique. Afin de permettre à l'utilisateur d'améliorer le résultat pour mieux répondre à ses préférences, notre système est interactif. Un objectif important de notre projet est le développement d'approches novatrices pour la communication d'une grande masse d'informations pertinentes à l'utilisateur tout en tenant compte des paramètres spécifiques à chaque utilisateur. (voir figure 1.5)

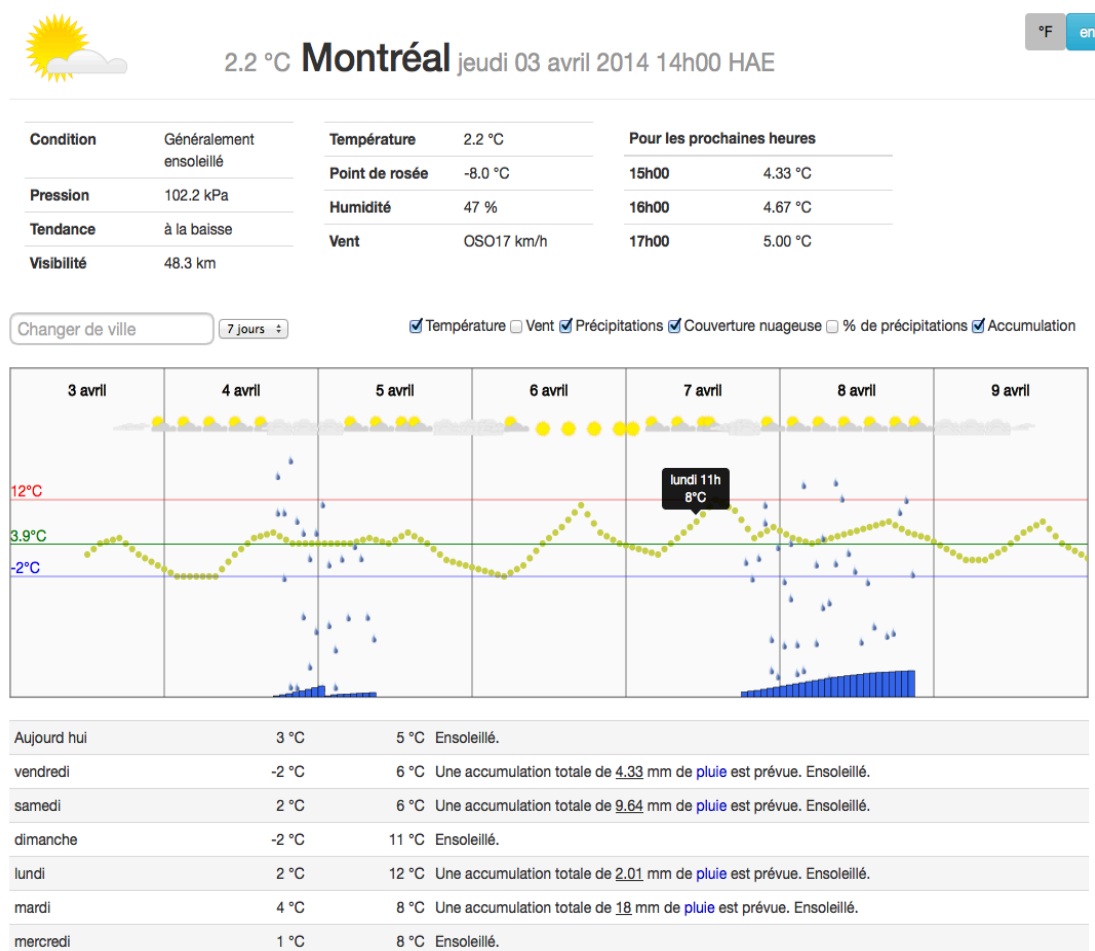


Figure 1.5 – Exemple de visualisation générée par MeteoVis

Un affichage sélectif de ces informations personnalisées permettrait à EC de fournir au public des prévisions mieux ciblées dans le temps et l'espace que celles produites actuellement en utilisant le fichier XML appelé Citypage (figure 1.3) qui est le résumé

du fichier MétéoCode.

1.6 Conclusion

En utilisant les méthodes et les approches de visualisation et en tenant compte de la quantité d'information environnementale que produit EC, l'affichage de toutes ces informations est impossible. La tâche principale de notre travail est de créer des méthodes et des approches novatrices permettant la visualisation d'une grande masse d'information. Cette présentation doit être personnalisée pour chaque usager. Nous appliquons ces méthodes dans notre domaine d'application. Nous avons créé un générateur de bulletins météorologiques qui générera chaque usager des bulletins personnalisés.

CHAPITRE 2

ÉTAT DE L'ART : TECHNIQUES, MÉTHODES ET APPROCHES DE VISUALISATION

L'objectif de cette thèse est d'élaborer des nouvelles méthodes et approches pour résoudre le problème de visualisation afin de permettre au public canadien de consulter toute l'information environnementale dont il a besoin.

La visualisation de l'information est devenue une nécessité et non un luxe. Plusieurs domaines l'utilisent pour simplifier l'affichage d'une grande masse d'information comme les résultats de la recherche web Nguyen et Zhang (2006), Wiza et al. (2004). Parmi les travaux antérieurs pertinents à notre recherche, nous retrouvons des travaux théoriques sur la conception de présentations visuelles, les techniques de visualisation, l'étude de la perception, des études psychologiques et des travaux dans le domaine de la vision. Dans ce chapitre, nous passerons en revue les travaux les plus pertinents dans ces domaines et nous soulignerons leurs points importants ainsi que leur influence sur notre projet.

Ma thèse poursuit les travaux de Fasciano (1996) qui a développé un modèle pour résoudre le problème de la génération de texte et de graphiques intégrés dans les rapports statistiques. Il a considéré plusieurs critères tels que l'intention du rédacteur, les types de variables, les relations entre ces variables et les valeurs des données. L'inspiration vient du fait que nous voulons aussi générer des bulletins contenant des graphiques et du texte. Un critère très important dans notre projet est le fait que notre visualisation doit être personnalisée et que l'utilisateur ait la possibilité de paramétrer le résultat et l'adapter mieux à ses besoins. Ce point sera étudié dans la section 2.4 qui traite de l'interactivité.

2.1 Génération des graphes

L’affichage des données est crucial pour leur analyse. La visualisation de ces données nous permet de mieux les explorer et de constater leurs tendances générales et leur comportement. La tâche de construction d’un graphe est l’encodage d’une quantité d’information catégorique et quantitative par le moyen de méthodes d’affichage. Cleveland (1994) étudie les principes de construction des graphes. Cette étude détermine les principes qui peuvent améliorer la capacité d’un graphique pour montrer la structure des données. Ces principes sont basés sur l’étude de la perception graphique. Cette étude concerne les 5 graphes les plus connus (voir figure 2.1) qui sont décrits dans Zelazny (2001). Ces graphes sont :

La tarte

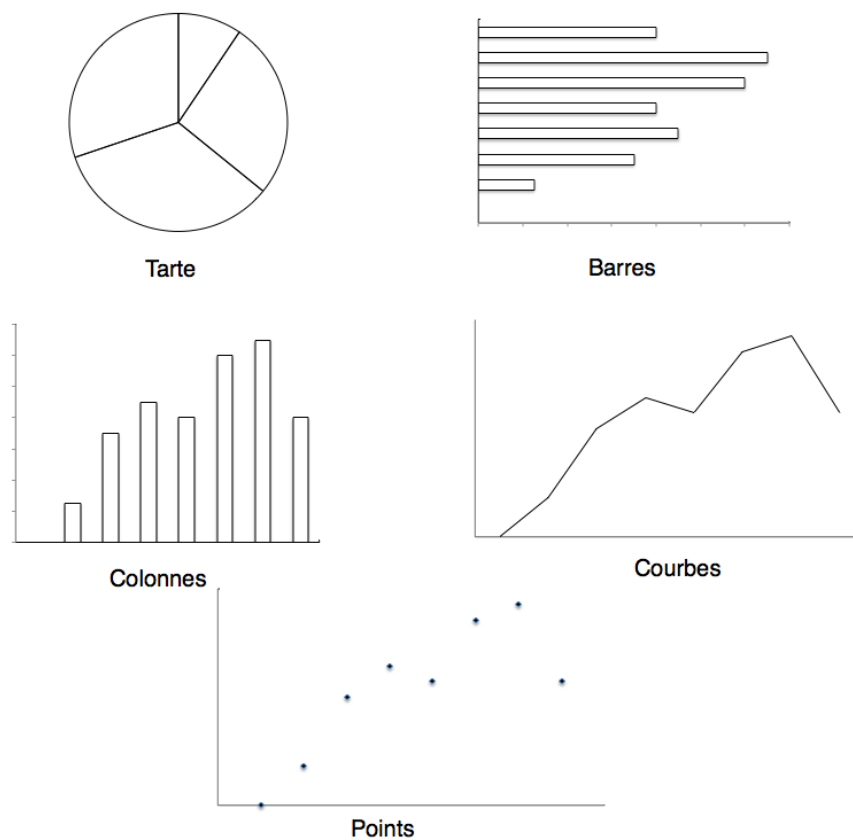
Son but principal est de montrer l’importance relative des composants l’un par rapport à l’autre et à l’ensemble. Pour faire comprendre une comparaison de décomposition, le mieux est d’utiliser une tarte. Une tarte est un diagramme circulaire constitué d’un cercle divisé en segments en forme de coin. La superficie de chaque segment (parfois appelée tranche) est en fonction du pourcentage du cercle total.

Les barres

Le graphique barre est celui qui convient le mieux pour démontrer une position. Dans un graphique à barres, la taille de chaque barre est proportionnelle à la valeur qu’il représente. Le graphique à barres n’a pas des marques de graduation. Le but principal d’un graphique à barres est d’orienter visuellement le spectateur à la taille relative des divers éléments d’une série de données avec une échelle quantitative sur l’axe horizontal. La dimension verticale n’est pas une échelle ; elle est entièrement consacrée aux intitulés des éléments mesurés.

Les colonnes

Figure 2.1 – Les 5 graphiques les plus utilisés



Les colonnes sont utilisées pour montrer l'évolution au fil du temps. Une évolution sera efficacement représentée au moyen soit d'une colonne, soit d'une courbe. La courbe est privilégiée si nous avons un grand nombre d'unités de temps (les jours d'une année). Les colonnes sont préférées si notre laps de temps est divisé en quelques parties (6 ou 7) par exemple les trimestres d'une année. Contrairement aux barres, les colonnes sont dans une position verticale.

La courbe

Ce graphique est le plus facile à élaborer, le plus souple et celui qui montre avec

le plus de clarté l'augmentation, la diminution, la fluctuation ou la stabilité d'une tendance.

Points

Ce graphe affiche des informations quantitatives par le biais de points de données représentés par point ou d'autres symboles. Il est fréquemment utilisé pour analyser les relations entre deux ou plusieurs variables. Il est considéré comme l'un des meilleurs types de graphiques pour enquêter sur la corrélation potentielle entre deux ensembles de données.

Ces graphiques sont les plus connus et les plus utilisés. Cependant, il en existe une bonne centaine d'autres. Harris (1999) est une référence dans le domaine des graphiques. Il peut être considéré comme étant un dictionnaire de graphes possibles. Dans ce dernier, nous distinguons quatre représentations graphiques utilisées dans le domaine de la météo.

Température : des bandes de couleur sont utilisées pour indiquer les zones de plages de température égale. Cet usage est parfois appelé une isoplèthe¹ (bandes de valeurs égales), ou une isotherme² (bandes d'égale température.). La gamme de valeur qui s'applique est généralement indiquée dans les bandes. Par exemple, une bande étiquetée 40 signifie que les températures pour les zones comprises dans la plage de 40 à 49 degrés

Pression : lorsque les lignes relient les points de même pression, les cartes est parfois appelé une carte isoligne ou isobare. Les valeurs sont indiquées sur les lignes et sont généralement exprimées en termes de millibars (mb), en utilisant uniquement les deux ou trois derniers chiffres. Par exemple, 20 sur la carte correspond à une valeur de 1020 mb 207 mb équivaut 1020,7 et 996 est égale à 999,6 mb. Les lettres

1. Une ligne reliant les points ayant les mêmes précipitations à la surface du globe.

2. Se dit d'une ligne reliant sur une carte des points où la température est identique à un moment donné.

H (High)et L (Low) désignent respectivement les points de pression le plus élevée et le plus bas dans la région.

Front³ chaud et froid : Symboles pour transmettre des informations sur le type des fronts météorologiques, où il se trouve, et les directions de son mouvement. six des symboles les plus couramment utilisés pour les fronts météorologiques sont présentés dans la figure 2.2.

Informations météorologiques locales : certaines cartes fournissent des données très détaillées pour un emplacement spécifique en utilisant une combinaison de symboles météorologiques et des données numériques. L'exemple de la figure 2.3 montre comment les symboles sont utilisés pour désigner la direction du vent. Plusieurs autres symboles sont représentés dans la figure 2.4. La plupart de ces symboles sont utilisés par les professionnels et ne seraient pas utiles pour le grand public.

L'utilisation des graphes est sans aucune doute essentielle pour résumer et communiquer l'information. La tâche de génération des graphes ne se résume aucunement en l'encodage d'une masse d'information sous forme graphique. Il faut tenir compte du décodage qui sera fait par l'utilisateur. Si l'utilisateur ne réussit pas à décoder l'information contenue dans le graphe, nous considérons que la génération du graphe a échoué. C'est pourquoi la génération de graphiques doit tenir compte de la perception humaine. Robbins (2012) explique comment créer la meilleure présentation en tenant compte de tous les paramètres (choix de type de graphique, la quantité d'information, choix d'attributs de style...) et de la perception vis à vis des graphes.

3. Un front météorologique est une surface de discontinuité étendue, qui sépare deux masses d'air ayant des propriétés physiques différentes (source : Wikipédia)

Figure 2.2 – Les fronts météorologiques

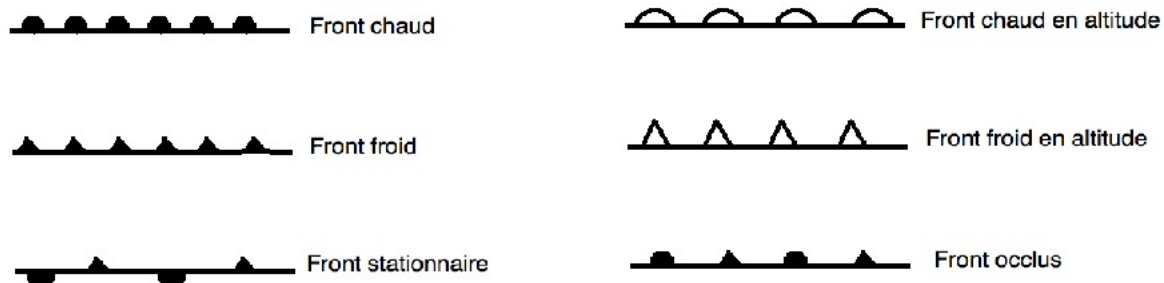


Figure 2.3 – Les symboles des directions du vent

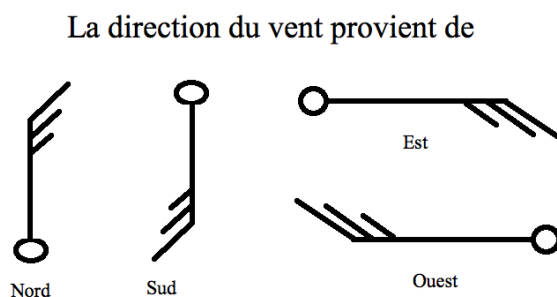


Figure 2.4 – Les symboles météo

Type of weather			Type of clouds	Amount of clouds
☉ Intermittent drizzle	* Snow shower	= Mist	☁ Altostratus	○ Clear sky
☉☉ Continuous drizzle	* Intermittent snow	≡ Fog	☁ Altostratus	① A few clouds
☉☉☉ Rain showers	* * Continuous snow		☁ Cirrocumulus	☉ Scattered clouds
☉☉☉☉ Intermittent rain	+ Slightly drifting snow	⚡ Lightning	☁ Cirrostratus	☉☉ Partly cloudy
☉☉☉☉☉ Continuous rain	+ Heavy drifting snow		☁ Cirrus	☉☉☉ Mostly cloudy
☉☉☉☉☉☉ Thunderstorm	△ Sleet		☁ Cumulonimbus	☉☉☉☉ Cloudy/overcast
☉☉☉☉☉☉☉ Heavy thunderstorm	▲ Hail shower		☁ Cumulus	⊗ Sky obscured
☉☉☉☉☉☉☉☉ Squall	☉☉ Freezing rain		☁ Nimbostratus	
☉☉☉☉☉☉☉☉☉ Tropical storm			☁ Stratocumulus	
☉☉☉☉☉☉☉☉☉☉ Hurricane			☁ Stratus	

2.2 Techniques de visualisation

L'objectif principal de toute visualisation est d'aider un utilisateur à analyser une grande quantité d'informations. Les nouvelles technologies nous permettent de générer des visualisations beaucoup plus complexes qu'un simple graphe et qui présentent beaucoup plus d'informations que peut contenir un graphe. Selon le type de l'information, nous pouvons choisir la technique qu'on peut utiliser. L'information peut avoir plusieurs formes (données unidimensionnelles, bidimensionnelles, multidimensionnelles, (les algorithmes et logiciels), etc). La visualisation peut être sous différentes formes en utilisant plusieurs techniques : standard 2D/3D, à base d'icônes, empilés, etc Keim (2002)

La sélection et la création de la conception la plus efficace parmi toutes les alternatives pour une situation donnée exigent habituellement beaucoup de connaissances et de créativité de la part du concepteur. La compréhension des caractéristiques des données ainsi que les propriétés graphiques pertinentes sont importantes dans la construction de techniques de visualisation. Mais aussi la compréhensibilité de toute image ou graphique est essentielle pour la présentation efficace de l'information inhérente dans les données. L'utilisateur doit être capable de déchiffrer l'information contenue dans la visualisation. Ce paramètre, qui est l'utilisateur, doit être pris en considération dans la création de la visualisation.

2.3 Perception

La visualisation des données est efficace lorsqu'elle établit un équilibre entre la perception et la cognition pour mieux tirer parti des capacités du cerveau. La perception visuelle, qui est gérée par le cortex visuel situé à l'arrière du cerveau, est extrêmement rapide et efficace. Nous voyons tout de suite, avec peu d'effort. La perception cognitive, qui est gérée principalement par le cortex cérébral à l'avant du cerveau, est beaucoup

plus lente et moins efficace. La construction du sens de données et des méthodes de présentation nécessite une réflexion consciente pour la presque totalité des travaux. La visualisation des données déplace l'équilibre vers une plus grande utilisation de la perception visuelle, en profitant de la puissance de nos yeux.

Pineo et Ware (2011) présentent un modèle de traitement de l'information de la perception visuelle humaine. Ce modèle est en deux étapes. Dans la première, l'information est traitée en parallèle pour extraire les caractéristiques de base de l'environnement. Dans la deuxième, l'attention visuelle joue un rôle beaucoup plus actif que dans la première étape et les éléments de l'environnement ont tendance à être examinés dans l'ordre.

Étape 1 : Processus parallèle pour extraire les propriétés de bas niveau de la scène visuelle

L'information visuelle est d'abord traitée par de grands réseaux de neurones dans les yeux et dans le cortex visuel primaire à l'arrière du cerveau. Les neurones individuels sont sélectivement stimulés à certains types de renseignements, tels que l'orientation des bords ou la couleur d'une tache de lumière. Dans chaque sous-zone, de grands réseaux de neurones fonctionnent en parallèle, extrayant des caractéristiques particulières de l'environnement. Au début, ce traitement parallèle est efficace, et il est largement indépendant de ce que nous choisissons de suivre. Il est également rapide. Si nous voulons que les gens comprennent vite l'information, nous devrions la présenter de telle manière qu'elle puisse être facilement détectée par ces grands systèmes de calcul rapide dans le cerveau.

Etape 2 : traitement séquentiel dirigé par les buts

À la deuxième étape, il y a une bifurcation en un sous-système spécialisé pour la reconnaissance d'objets et un sous-système spécialisé pour interagir avec l'environnement. Dans le cas de la reconnaissance d'objets, des facteurs tels que l'attention visuelle et la mémoire deviennent importants. De toute évidence, pour identifier un objet, les spectateurs doivent, en quelque sorte, faire correspondre les caractéristiques visuelles avec des

propriétés de l'objet stocké dans la mémoire. En général, les tâches que l'observateur effectue influent sur ce qui est perçu. L'un des principaux mécanismes concernant ce qui est perçu à une partie précise de la visualisation est l'attention visuelle. Nous savons que certains aspects de cette deuxième transformation se produisent de manière séquentielle, un seul objet visuel est traité à la fois.

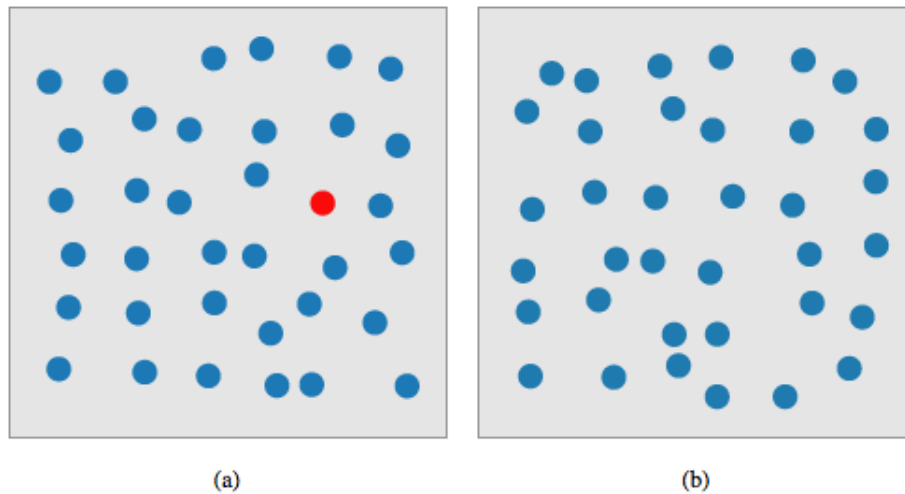
2.3.1 Traitement préattentif

Pendant de nombreuses années, les chercheurs ont étudié comment le système visuel humain analyse les images. Un résultat important a été la découverte initiale d'un ensemble limité de propriétés visuelles qui sont détectées très rapidement et avec précision par le système de bas niveau visuel. Ces propriétés ont été initialement appelées préattentives, car leur détection semblait précéder une attention particulière. Nous savons maintenant que l'attention joue un rôle essentiel dans ce que nous voyons, même à ce stade précoce de la vision. Cependant, le terme préattentive continue d'être utilisé, car il exprime une notion intuitive de la vitesse et la facilité avec laquelle ces propriétés sont identifiées.

Un exemple simple d'une tâche préattentive est la détection d'un cercle rouge dans un groupe de cercles bleus (figure 2.5). L'objet cible a un visuel de propriété "rouge" et les objets distracteurs sont bleus (tous les objets non-cibles sont considérés comme distracteurs). Un spectateur peut dire à un coup d'œil si la cible est présente ou absente.

Dans notre domaine d'application, ce concept peut nous aider à mettre en évidence les avertissements concernant les intempéries. Cela permettrait aux usagers de détecter immédiatement les mauvaises conditions météorologiques. Les avertissements dans le domaine météorologique sont d'une grande importance.

Figure 2.5 – Un exemple de recherche d'un cercle cible rouge basée sur une différence de couleur : (a) la cible est présente dans une mer de cercle distracteurs bleu (b) la cible est absente



2.3.2 Les principes de Gestalt

Gestalt est un terme de psychologie qui signifie « tout unifié ». Il se réfère aux théories de la perception visuelle développée par des psychologues allemands dans les années 1920. Ces théories tentent d'expliquer comment les gens ont tendance à organiser les éléments visuels en groupes ou ensembles unifiés où certains principes sont appliqués. Les lois de Gestalt se traduisent facilement en un ensemble de principes de conception pour les visualisations d'information. Les lois de Gestalt et des principes de conception connexes sont présentés ci-dessous Ware (2012).

La proximité : La proximité spatiale est un principe d'organisation et de perception.

Les objets proches les uns des autres sont perceptivement regroupés. La loi de proximité suppose que lorsque nous percevons une collection d'objets, nous allons voir les objets proches comme formant un groupe.

La similitude : Les formes des éléments peuvent aussi déterminer la manière dont ils sont regroupés. La loi de similitude saisit l'idée que les éléments seront regroupés

perceptuellement s'ils sont semblables.

La continuité : Le principe de continuité indique que nous sommes plus enclins à construire des entités visuelles à partir d'éléments visuels lisses et continues, plutôt que ceux qui contiennent de brusques changements de direction. Il devrait être plus facile d'identifier les points de départ et d'arrivée de lignes de connexion si elles sont lisses et continues.

La symétrie : La loi de symétrie adopte l'idée que lorsque nous percevons des objets, nous avons tendance à les percevoir comme des formes symétriques établies autour de leur centre. La plupart des objets peuvent être divisés en deux moitiés plus ou moins symétriques et quand, par exemple, nous voyons deux éléments sans lien qui sont symétriques, inconsciemment nous les intégrons dans un objet cohérent (ou percept). Plus les objets sont semblables, plus ils ont tendance à être regroupés.

Le contour : Un contour fermé tend à être considéré comme un objet. Partout où un contour fermé est vu, il y a une tendance très forte de la perception à diviser les régions de l'espace en « intérieur » et « extérieur » du contour. Beaucoup de contours fermés sont utilisés pour délimiter les relations entre différents ensembles qui se chevauchent.

Dans notre analyse du site actuel d'environnement Canada et des sites de ses concurrents (section 2.8) nous ne pouvons pas parler de ces principes proprement dits. La principale raison de cela est que les graphiques utilisés se limitent à des icônes génériques montrant le résumé de l'état météorologique (soleil, nuages, pluie, neige...) d'une période de temps bien défini (une journée, une période de la journée ou une heure précise). Le seul principe utilisé est le contour qui est utilisé pour limiter une zone bien définie contenant des conditions météorologiques identiques ou similaires.

2.4 Interactivité

En plus des techniques de visualisation, pour une exploration de données efficace, il est nécessaire d'utiliser une certaine interaction. Les techniques d'interaction permettent à l'analyste de données d'interagir directement avec les visualisations et de changer dynamiquement les visualisations en fonction des objectifs d'exploration, et ils permettent aussi de relier et de combiner des visualisations indépendantes. Dans le domaine de la météo, il y a plusieurs détails qu'on peut afficher pour chaque paramètre de la visualisation. Toutefois, il est impossible de toutes les afficher en mêmes temps. L'interactivité est une bonne solution pour que l'utilisateur puisse adapter la visualisation à ses besoins et préférences. Keim (2002) propose une classification des techniques de visualisation de l'information et de data mining basée sur le type de données à visualiser, la technique de la visualisation et la technique d'interaction et de déformation.

La motivation pour l'interaction est claire, mais nous devons tenir compte de ce qui motive l'utilisateur à interagir. Yi et al. (2007) ont réalisé une étude pour répondre à cette question. Ils se sont basés sur différentes intentions des usagers et ont introduit une liste de catégories qui décrivent pourquoi les utilisateurs souhaiteraient interagir. Dans la suite, nous utilisons les catégories de Yi et al. (2007) :

Selectionner - Marquer quelque chose d'intéressant

Lorsque l'utilisateur repère une partie intéressante dans la représentation visuelle, il veut la marquer et la mettre en évidence en tant que telle, que ce soit temporairement pour des résultats intrigants ou de façon permanente pour mémoriser des résultats importants.

Explorer - Montrez-moi autre chose

Pour que la visualisation d'une masse complexe de données qui varie dans le temps soit pratique, l'utilisateur doit se concentrer sur une plage de temps et sur une partie

des variables de données. En conséquence, les usagers doivent être capables de consulter de manière interactive les différentes parties du domaine du temps et pouvoir considérer les variables alternatives pour l'inclusion dans le codage visuel pour arriver à une vue globale des données.

Reorganiser - Montrez-moi un arrangement différent

Différentes organisations possibles de temps et des données associées peuvent communiquer des aspects complètement différents, un fait qui devient évident se rappelant la distinction entre les représentations linéaires et cycliques du temps. Comme les usagers veulent regarder le temps sous des angles différents, ils doivent être pourvus d'installations qui leur permettent de générer interactivement différentes dispositions spatio-temporelles orientées données.

Encoder - Montrez-moi une représentation différente

De même pour ce qui a été dit à propos de la disposition spatiale, le codage visuel de valeurs des données a un impact majeur sur ce qui peut être dérivé d'une représentation visuelle. Parce que les données et les tâches sont variées, les utilisateurs doivent être capables d'adapter le codage visuel pour répondre à leurs besoins, que ce soit pour effectuer des tâches de localisation ou de comparaison, ou pour confirmer une hypothèse générée à partir d'un encodage visuel en le vérifiant avec une autre alternative.

Résumer / Elaborer - Montrez-moi plus ou moins de détails

Lors de l'analyse visuelle, les usagers ont besoin de regarder certaines choses en détail, tandis que pour d'autres des représentations schématiques sont suffisantes. Les niveaux de granularité structurés hiérarchiquement, où les abstractions de haut niveau fournissent des aperçus agrégés, et les niveaux inférieurs, les détails correspondants.

Filtrer - Montrez-moi quelque chose de conditionnel

Lorsque les usagers recherchent des informations particulières dans les données ou évaluent une certaine hypothèse sur les données, il est logique de restreindre la visualisation pour n'afficher que les éléments qui respectent les conditions imposées par les critères de recherche ou les contraintes de l'hypothèse. En filtrant interactivement ou en atténuant les éléments de données non pertinentes, nous éclaircissons la visualisation pour les usagers et leur permettant de se concentrer sur leur tâche en cours.

Connecter - Montrez-moi les éléments liés

Lorsque les utilisateurs font une découverte potentiellement intéressante dans les données, ils se demandent généralement si des découvertes similaires ou connexes peuvent être faites dans d'autres parties des données. Ainsi, les utilisateurs ont l'intention, de façon interactive, de trouver, comparer et évaluer de telles similitudes ou des relations, par exemple, pour voir si une tendance qu'ils ont découverte dans une saison d'une année est présente pour les variables d'autres données ou se répète dans le même temps dans les années subséquentes.

Annuler / Refaire - Laissez-moi aller où j'ai déjà été

Les utilisateurs ont à naviguer dans le temps et regarder à différents niveaux de granularité, ils doivent essayer différentes modalités et codages visuels, et ils ont à expérimenter avec des conditions de filtrage et de seuils de similarité. Pour tenir compte de la nature exploratoire et interactive du raisonnement analytique, un mécanisme d'historique avec les annuler et refaire des opérations sont nécessaires. Annuler / refaire permet aux utilisateurs d'essayer de nouvelles vues sur les données et de retourner sans effort à la représentation visuelle précédente si la nouvelle n'a pas fonctionné comme prévu.

Changer la configuration - Permettez-moi d'ajuster l'interface

En plus d'adapter la représentation visuelle des données et des tâches à accom-

plir, les utilisateurs veulent également adapter le système d'ensemble qui assure la visualisation. Cela inclut l'adaptation de l'interface utilisateur (par exemple, l'arrangement des fenêtres ou les éléments dans les barres d'outils), mais aussi la gestion générale des ressources du système (par exemple, la quantité de mémoire utilisée).

Dans leur ensemble, ces intentions constituent ce qu'un système de visualisation doit soutenir en termes d'interaction, afin de profiter pleinement de la synergie des humains et des capacités de la machine.

Alors que le marquage (ou la sélection) des éléments intéressants et la navigation dans le temps sont quasi obligatoires, les installations pour d'autres intentions ne sont pas souvent parvenues à un état de développement suffisant ou ne sont pas encore prises en considération. Ceci est probablement dû à l'effort supplémentaire que nous devons dépenser pour la mise en œuvre de méthodes d'interaction efficaces. Mais en fait, toutes ces intentions des usagers sont toutes aussi importantes et les techniques correspondantes devraient être fournies.

2.5 Génération de texte

La génération du langage naturel (NLG) est un champ de l'intelligence artificielle et de la linguistique computationnelle qui traite des systèmes de construction de logiciels informatiques qui peuvent produire des textes significatifs en langue naturelle à partir d'une représentation sous-jacente non linguistique de l'information. Les systèmes NLG utilisent des connaissances sur la langue et le domaine d'application pour produire automatiquement les documents, rapports, messages d'aide, et d'autres types de textes.

SumTime (Sripada et al., 2001) est un modèle générique de calcul pour la production de résumés textuels à partir d'une série chronologique de données. SumTime est constitué d'un modèle en deux étapes pour la détermination du contenu. La première

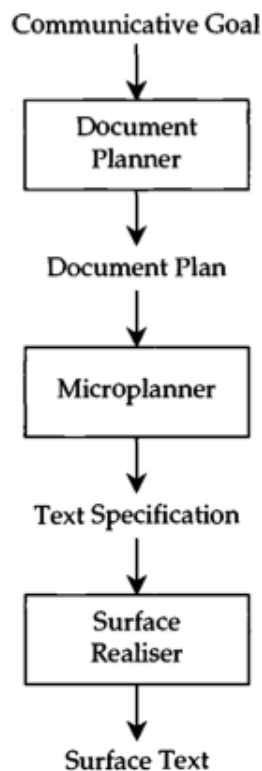
étape consiste à construire une vision qualitative de l'ensemble des données, et la seconde consiste à utiliser cet aperçu, avec les données réelles, afin de produire des résumés. SumTime a été testé sur les prévisions météorologiques (Sripada et al., 2003) et le diagnostic des turbines à gaz (Yu, 2004). Cette approche est inspirée de la méthodologie décrite dans Reiter et Dale (2000) qui décrit le processus de génération du langage naturel. Cette méthodologie (voir figure 2.6) est basée sur une décomposition architecturale particulière du processus de la génération du langage naturel en trois modules : la planification du document, la microplanification, et la réalisation de la surface. La planification du document est ce qui est souvent appelé « la planification du texte », et comprend deux sous-tâches : la détermination du contenu et la structuration du document. La microplanification comprend l'agrégation, la génération d'expressions référentielles, et certains aspects de lexicalisation. La réalisation de la surface comprend la réalisation linguistique et la réalisation de structure.

En termes de représentations intermédiaires, la sortie du planificateur de document dans ce modèle est une spécification de document. Ceci est un arbre constitué d'informations portant des unités appelées messages, souvent avec des relations de discours spécifiées entre les parties de l'arbre. La sortie de la microplanification est une spécification du texte. C'est un arbre dont les feuilles spécifient les caractéristiques de la phrase et dont les nœuds internes spécifient la structure logique du document en termes de paragraphes, sections, etc.

2.6 Génération combinée des textes et des graphiques

Un graphe simplifie beaucoup la tâche de l'utilisateur pour analyser une grande quantité d'information. Toutefois, à lui seul, il peut être difficile de décoder tout le contenu. Le texte et les graphiques jouent des rôles complémentaires dans la transmission de l'information à l'utilisateur. Fasciano (1996) décrit le système PostGraphe qui génère des rapports

Figure 2.6 – Architecture d'un système de génération de langue naturel (Reiter et Dale, 2000)



statistiques contenant du texte et des graphiques en se servant d'une description annotée des données à présenter. Les annotations utilisées correspondent aux critères établis dans le modèle théorique. Ainsi, l'utilisateur peut spécifier au système ses intentions (comparaison, évolution, réparation, corrélation...), les types des données à présenter (temporelles, numériques, ordonnées, . . .) et les relations entre les données.

SelTex (Corio et Lapalme, 1999), est la suite des travaux de Fasciano (1996). C'est un système de génération qui produit des textes courts et des légendes pour accompagner les graphes qui sont générés selon les intentions de l'auteur. SelTex utilise des règles qui ont été extraites d'une étude de corpus de plus de 400 extraits de texte. Ces règles sont à la base de la génération d'un texte simple qui décrit les tendances générales des données des graphes (évolution, comparaison, corrélation...). D'une manière similaire,

mais moins simple, Mittal et al. (1998) décrit un système pour produire des légendes pour les tableaux complexes. Ce système détermine le contenu et la structure du sous-titrage en analysant la structure des représentations graphiques et de la complexité de ses éléments perceptifs et en utilisant des transformations linguistiques telles que la commande, l'agrégation et le centrage.

2.7 Analyse de l'existant

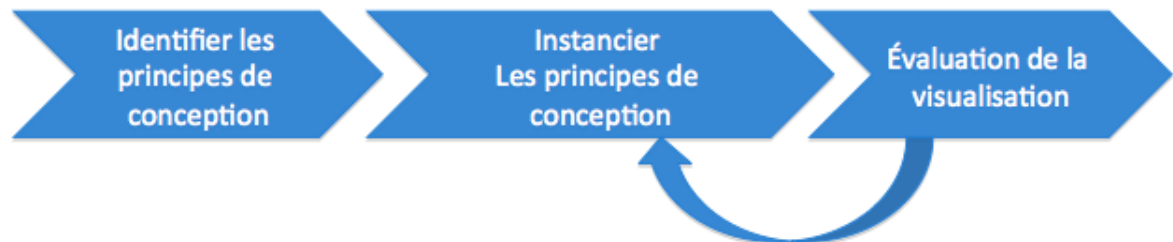
Après cette revue des travaux en rapport avec la génération de graphe, les techniques de visualisation, la perception, l'interactivité, la génération de texte et la génération combiné de texte et de graphique, nous étudions la création de la visualisation.

Pour créer une visualisation, Lam et al. (2012) proposent de commencer par la compréhension de l'environnement et des pratiques de travail p.e. l'observation de terrain, des entretiens et l'observation en laboratoire. Agrawala et al. (2011) suggèrent comme première étape de leur approche pour la création d'une visualisation, d'explorer des visualisations existantes dans le même domaine pour identifier des principes de conception. Nous avons choisi de suivre ces étapes (voir Figure 2.7) : la première est l'analyse des visualisations conçues dans le même domaine. Nous appliquons cette étape en analysant le site actuel d'Environnement Canada et de ses concurrents ainsi que d'autres sites internationaux. Les sites que nous analysons pour cette première étape sont :

- Météo Canada (figure 2.9 le site actuel d'Environnement Canada).
- MétéoMédia (Un site privé canadien de la météo).
- NOAA (Le site gouvernemental de la météo au États-unis).
- Météo France (Le site gouvernemental de la météo en France).
- Météo Blue (Un site privé suisse de la météo mondiale).

Après notre analyse de ces visualisations conçues à la main, nous avons essayé d'ex-

Figure 2.7 – Les étapes de l’approche de Agrawala et al. (2011) pour la conception d’une visualisation



traire et de dégager des règles et des principes pour la visualisation de la météo. Parmi les principes déduits, nous pouvons citer que l’affichage des avertissements est fondamental pour un système d’information météorologique. Un autre principe qui pourra faciliter l’interprétation de notre visualisation est la combinaison du texte et des graphiques. Le principe que nous considérons comme le plus important est la personnalisation de la visualisation pour chaque usager en tenant compte de son profil qui doit être détecté automatiquement et sans atteinte à sa vie privée et à ses informations personnelles. Pour cette raison, nous n’utiliserons pas de témoin de connexion (cookies). Nous pouvons toutefois nous baser sur plusieurs astuces pour profiler notre usager. Par exemple, nous pouvons estimer son emplacement (région) à l’aide de l’adresse IP utilisée pour consulter le site. Nous choisissons la langue d’affichage des messages en détectant la langue qu’il utilise dans son navigateur. Nous pouvons nous baser aussi sur son comportement sur le site selon l’étude faite dans Cadez et al. (2000).

Pour la deuxième étape, nous appliquons les principes de conception identifiés à la première étape. Nous prenons comme point de départ des expériences menées par des membres du RALI pour explorer le type d’information disponible chez EC, des prototypes web avaient été développés pour afficher l’information météorologique graphiquement en utilisant Protovis basé sur *Scalar Vector Graphics* (SVG) et un autre en utilisant des informations alphanumériques, mais placées géographiquement à l’aide

de Google Maps. Une troisième expérience a été réalisée en utilisant jqPlot. Ce dernier est un plugin jQuery pour créer des graphiques. Ces expérimentations offrent de nombreuses représentations ainsi qu'une interactivité poussée avec l'utilisateur. La figure 2.8⁴ montre un exemple de résultat des expérimentations réalisées au RALI. Dans cette visualisation, nous trouvons les conditions météorologiques et les prévisions pour chaque heure de la semaine. Nous pouvons aussi choisir l'emplacement directement sur une carte ou en indiquant le nom de la ville.

Cela a permis d'expérimenter différentes manières de combiner les informations publiées quotidiennement par Environnement Canada avec d'autres approches basées sur le Web. Bien que ces prototypes n'aient été pas mis en production, ils ont montré la possibilité d'intégrer l'information environnementale avec les applications Web de sorte qu'elle devienne plus accessible et utile.

La troisième étape de l'approche de Agrawala et al. (2011) est l'évaluation des visualisations conçues en se basant sur les principes obtenus suite à la première étape. Nous mesurerons la rétroaction qualitative des usagers avec des interviews et la rétroaction quantitative à l'aide des statistiques d'utilisation. Nous réaliserons des études plus formelles des usagers afin de vérifier dans quelle mesure nos visualisations améliorent l'interprétation de l'information. Les critères d'évaluation doivent quantifier l'efficacité de certains aspects de la visualisation. Notre visualisation doit être expressive et donc présenter toute l'information dont nous avons besoin sans causer un encombrement visuel pour qu'elle soit efficace et interprétable avec précision et rapidité.

2.8 La théorie confrontée à la pratique

Dans cette section, nous analysons les visualisations des conditions météorologiques existantes des sites (annoncées dans la section 2.7) par rapport aux techniques vues dans

4. Ce site web permettant la visualisation des informations météorologiques a été conçu par Alessandro Sordani pour le projet météo en collaboration avec Environnement Canada

Figure 2.8 – Visualisation des informations météorologiques (RALI)



ce chapitre pour en dégager les qualités et les insuffisances.

Dans cette thèse, nous nous concentrons sur l'utilisateur. Par conséquent, dans l'analyse des visualisations existantes nous devons nous positionner comme étant des utilisateurs pour pouvoir prendre en considération le point de vue utilisateur. Chen (2005) a étudié 10 problèmes non résolus dans ce domaine, les trois premiers étant les problèmes dans une perspective centrée sur l'utilisateur :

- la facilité d'utilisation : les informations contenues dans la visualisation devraient répondre aux besoins des utilisateurs ; On cite ici quelques études d'utilisabilité et d'évaluations de visualisation Elmqvist et Yi (2012), Plaisant (2004), Goldberg et Helfman (2011), Weaver (2010) ;
- perception : les principes de perception sont incorporés par des algorithmes pour

optimiser le calcul et produire une visualisation idéale d'un point de vue humain et non du point de vue machine McNamara et al. (2011) ;

- connaissance préalable de l'utilisateur : peut être considéré comme un paramètre du second problème. En général, les utilisateurs ont besoin de deux types de connaissances avant de comprendre le message transmis dans les informations affichées : comment utiliser un système d'affichage d'informations et la façon d'interpréter son contenu.

2.8.1 Environnement Canada

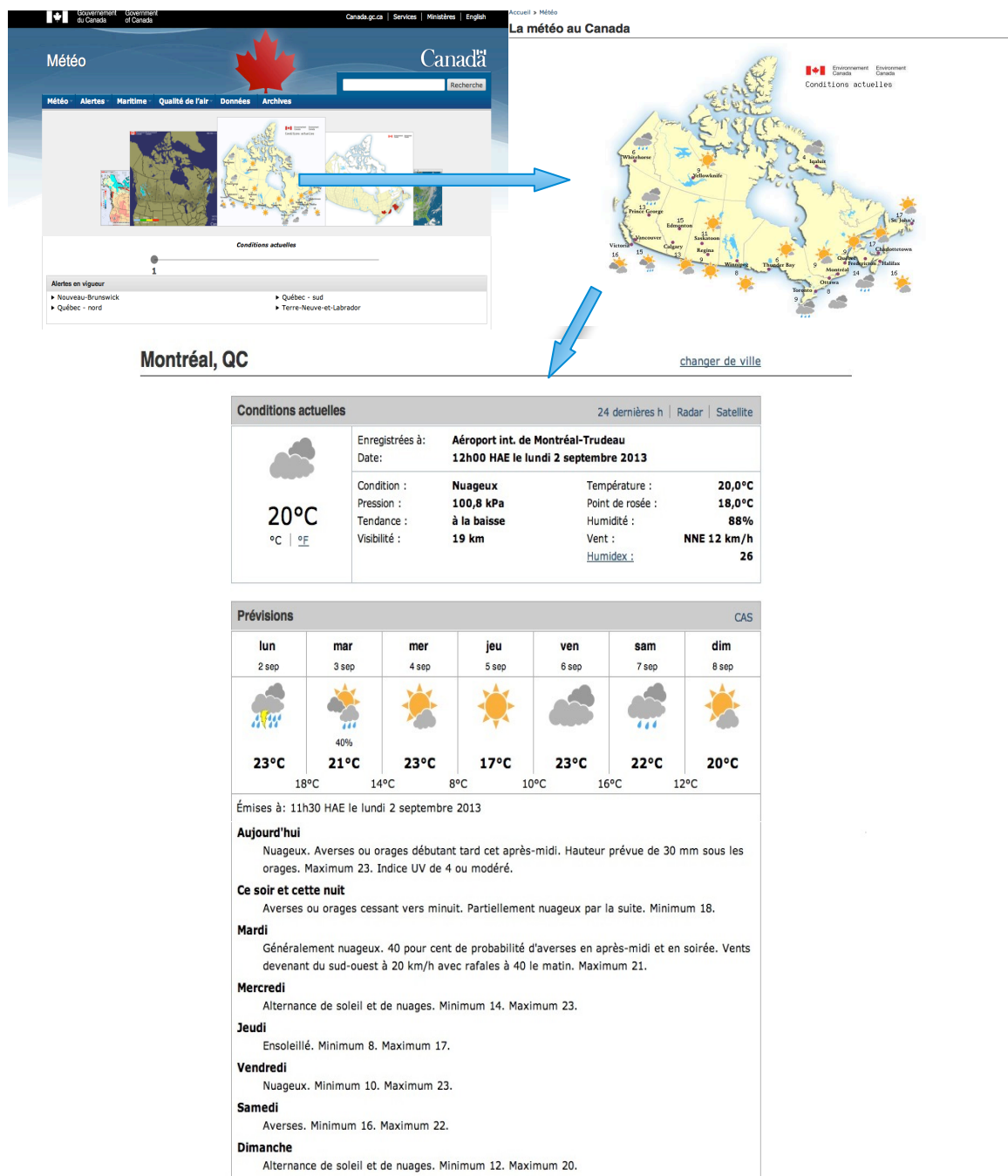
La page d'accueil du site d'Environnement Canada présente une "carte de prévision" cliquable qui résume les informations génériques sur les conditions météorologiques dans le pays. En cliquant sur la carte, l'utilisateur est redirigé vers la page de prévisions qui montre des prévisions textuelles et iconiques des prochains 7 jours pour cette région (figure 2.9).

La description des prévisions est à un trop gros niveau de granularité pour un utilisateur ayant des plans à court terme et voulant vérifier les informations météorologiques pour une heure précise de la journée. Les textes de prévisions de la semaine apparaissent ensemble, sans donner la possibilité à l'utilisateur de sélectionner et d'afficher uniquement les informations qui l'intéressent. L'interface utilisateur Web n'est pas interactive et aucune possibilité de personnalisation n'est offerte.

Les techniques de visualisation, qui visent la perception de l'utilisateur, étudiées dans la section 2.3 sont à peine utilisées dans les présentations générées par EC. Parmi toutes ces techniques, EC en utilise quelques-unes pour aider l'utilisateur à mieux percevoir la visualisation. Le principe de contour, utilisé dans la carte du Canada, pour délimiter les provinces et les territoires. L'interaction, à laquelle nous accordons une grande importance dans notre étude, est quasi absente dans les visualisations générées par EC.

L'exploration des données est offerte. L'utilisateur, de la page d'accueil, peut cliquer sur une région de la carte pour explorer les informations météorologiques de cette région. Aucune autre technique d'interaction est offerte.

Figure 2.9 – Le site officiel d'Environnement Canada : <http://www.meteo.gc.ca>. Pour arriver à cette page (une ville précise), l'utilisateur doit sélectionner la langue (français ou anglais) puis sélectionner la ville de son choix sur une carte ou dans une liste.



2.8.2 NOAA

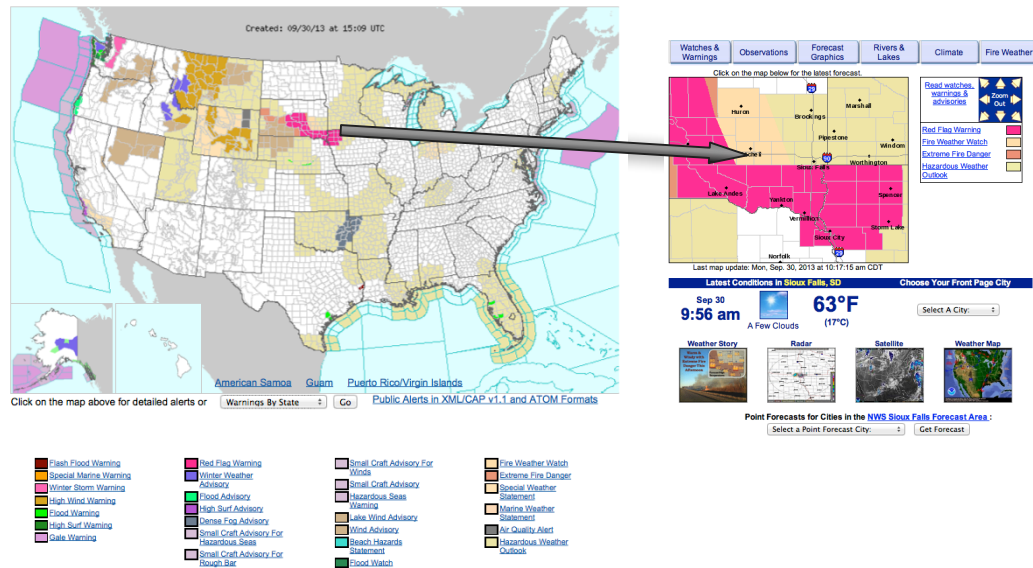
Le site de NOAA ⁵ (National Oceanic and Atmospheric Administration) est l'agence des États-Unis responsable de l'étude de l'océan et de l'atmosphère, c'est donc le site gouvernemental équivalent à celui d'EC pour le Canada. Le site web de NOAA affiche des avertissements à la page d'accueil. NOAA a mis au point des techniques graphiques permettant à l'utilisateur d'avoir facilement un aperçu de la situation nationale. Les visualisations de NOAA sont automatiquement mises à jour avec de nouveaux bulletins. Ils sont interactifs. Nous pouvons zoomer sur les États (voir figure 2.10). Un autre clic nous amènera à la page des prévisions détaillées. Les avertissements sont regroupés par catégories identifiées par des couleurs. Malgré l'exhaustivité des informations affichées, nous avons constaté que le grand nombre de catégories (une quarantaine) est ergonomiquement désavantageux. Pour savoir ce que signifie une couleur, nous devons consulter la légende. Mais le grand nombre de catégories ne permet pas un choix différencié des nuances des couleurs, ainsi plusieurs catégories auront des nuances de couleurs proches. L'utilisateur peut donc avoir des difficultés à distinguer les significations de chaque couleur.

Le système de prévision détaillée NOAA (voir figure 2.11) est basé sur la bibliothèque de Google Maps et génère une prévision précise pour des coordonnées spécifiques sur la carte. Un simple clic suffit pour régénérer les prévisions pour un nouvel emplacement. Les prévisions sont indiquées dans les deux modes textuels et graphiques utilisant des icônes sémantiquement riches. En bas de page de la prévision détaillée, une carte géographique indique une zone rouge, c'est cette zone qui est concernée par la prévision. La méthode utilisée dans NOAA pour indiquer les points de prévision détaillés est l'interpolation sur une grille.

NOAA fait un usage intensif du système de prévision graphique interactive. Une

5. <http://www.weather.gov>

Figure 2.10 – NOAA : visualisation et zoom des avertissements



grande quantité de cartes sont générées automatiquement par interpolation des valeurs environnementales comme la température, la vitesse du vent, du ciel couvert, la quantité de neige,...

Un grand nombre de visualisations sont générées par NOAA. Ces graphiques couvrent la quasi-totalité des informations météorologiques collectées et archivées. NOAA utilise plusieurs principes de perception et d'interactivité. Nous pouvons constater l'utilisation du principe de contour pour délimiter les états. Dans la page d'accueil, l'utilisation du principe de la similitude permet aux utilisateurs d'avoir une idée générale en un coup d'oeil (couleur similaire = condition similaire). L'interactivité est une technique bien présente. Elle est offerte aux utilisateurs de NOAA en leur permettant d'explorer plus de détails pour une région choisie en cliquant sur cette région. Aussi, l'usager peut choisir un arrangement des données qui lui plait en générant d'autres visualisations ou même encoder les mêmes informations d'une façon complètement différente. La technique de résumer/élaborer les données est partiellement utilisée. L'utilisateur peut consulter, dans des visualisations différentes, une visualisation pour chaque détail et une visualisation

Figure 2.11 – NOAA : Prévisions détaillées



pour tous les détails. Néanmoins, nous ne pouvons pas paramétrer des détails au choix pour les afficher dans la même présentation.

Des dizaines de présentations graphiques des informations météorologiques sont générées sur le site web de NOAA. L'inconvénient majeur est que ces présentations ne sont pas accompagnées d'une description textuelle.

2.8.3 MétéoFrance

Meteo France, équivalent français d'EC, distingue quatre catégories de mises en garde (rouge, orange, jaune et vert), sans préciser la nature exacte de ces avertissements qui sont signalés dans une boîte dans la page d'accueil⁶ (voir figure 2.12). En suivant

6. <http://france.meteofrance.com/>

ce lien, nous trouvons une carte géographique de couleur signalant des niveaux d'alertes faciles à comprendre. Nous pouvons alors cliquer sur chaque région pour obtenir un bulletin d'alerte plus précis.

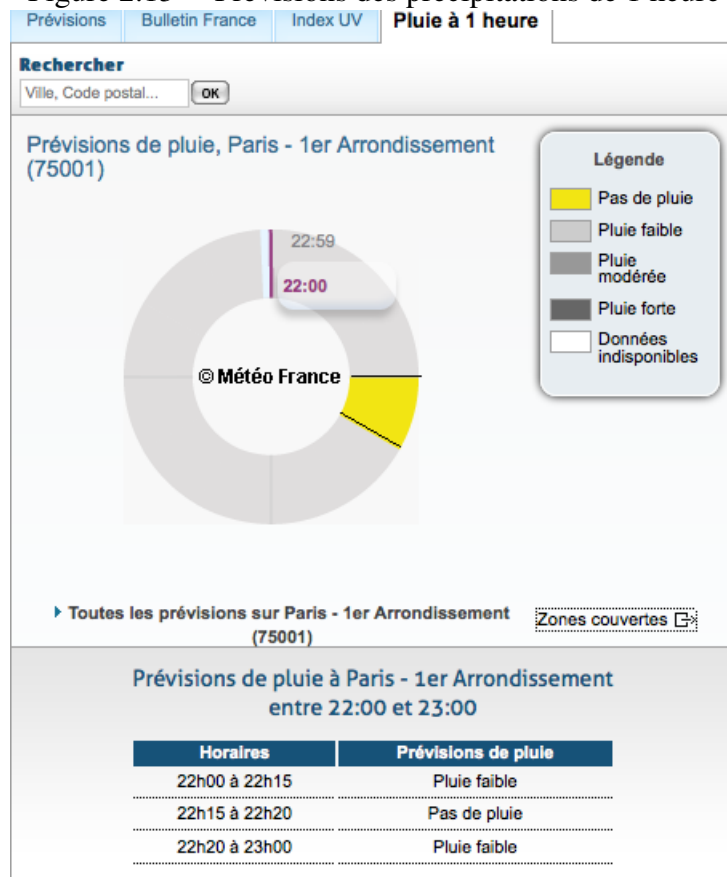
Figure 2.12 – Page d'accueil de Météo France



Nous avons trouvé intéressants certains systèmes d'affichage innovants déployés par les cas étudiés. Météo France a construit un outil visuel pour indiquer les prévisions de précipitations de 1 heure (voir figure 2.13).

La France est un pays beaucoup moins étendu que le Canada et les É-U. La masse d'information météorologique est donc moins importante. Ce qui laisse penser que les visualisations générées par météo France comportent plus de détails et de clarté. Ce qui

Figure 2.13 – Prévisions des précipitations de 1 heure



n'est pas du tout le cas. Météo France se contente d'afficher les informations qu'elle juge essentielles sans donner aux utilisateurs la possibilité d'interagir avec la visualisation pour demander plus de détails.

Le principe de contour est aussi présent dans météo France pour délimiter les villes. Nous pouvons explorer les détails d'une ville en positionnant le curseur sur cette région (température minimum et maximum et couverture nuageuse) ou bien, pour plus de détails, cliquer sur la région. Une très bonne description textuelle est générée dans l'onglet bulletin France. Elle décrit, textuellement, la condition météorologique globale. Météo France ne génère pas d'autre description textuelle du moins pour accompagner les graphiques.

2.8.4 Meteoblue

Meteoblue⁷ est un site privé présentant les conditions météorologiques partout dans le monde. Il détecte automatiquement l'emplacement de l'utilisateur à l'aide de son adresse IP et affiche par conséquent les conditions météo dans sa région. Meteoblue offre à l'utilisateur la possibilité de choisir un autre emplacement. Dans la page d'accueil, Meteoblue génère deux présentations :

- La première (figure 2.14) d'une apparence simple utilise des caractéristiques innovantes comme : (a) les pictogrammes pour indiquer la température, la vitesse et la direction du vent et l'indice UV. (b) Le rainSPOT pour donner un aperçu des précipitations dans un rayon de 35 km par un gradient de couleur. Cette visualisation présente les informations des six prochains jours. Pour explorer des détails plus fins pour une journée précise, l'utilisateur, en cliquant sur la journée, est redirigé vers une visualisation présentant les conditions pour chaque plage de trois heures de la journée.
- La deuxième (figure 2.15) d'une apparence moins simple que la première, mais indiquant plus de détails d'une manière plus technique en utilisant les symboles météorologiques vus dans les figures 2.3 et 2.4.

Dans cette visualisation plusieurs techniques sont utilisées : la continuité de la ligne indiquant la température montre l'évolution de celle-ci au fil du temps. La nuance du gris utilisée montre la similitude de la couverture nuageuse. Cependant, l'interactivité n'est pas le point fort de Meteoblue. Elle n'est présente que dans la première visualisation dans le fait d'explorer plus de détails et avoir une présentation (en bloc de trois heures) d'une journée.

7. <http://www.meteoblue.com>

Figure 2.14 – Visualisation générée par Meteoblue utilisant des pictogrammes pour indiquer la couverture nuageuse, les températures max et min, la vitesse et la direction du vent, l'indice UV et le rainSPOT pour indiquer la précipitation dans un rayon de 35 km.

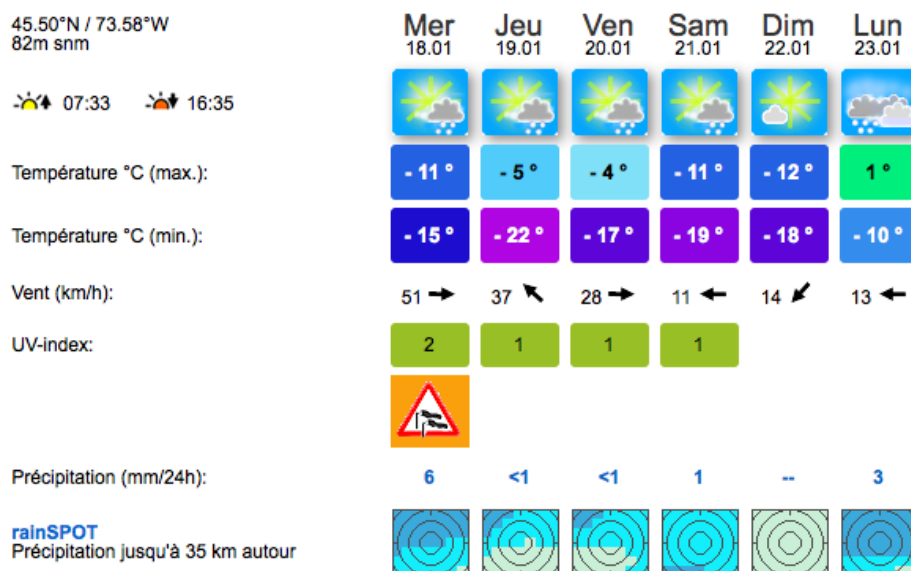
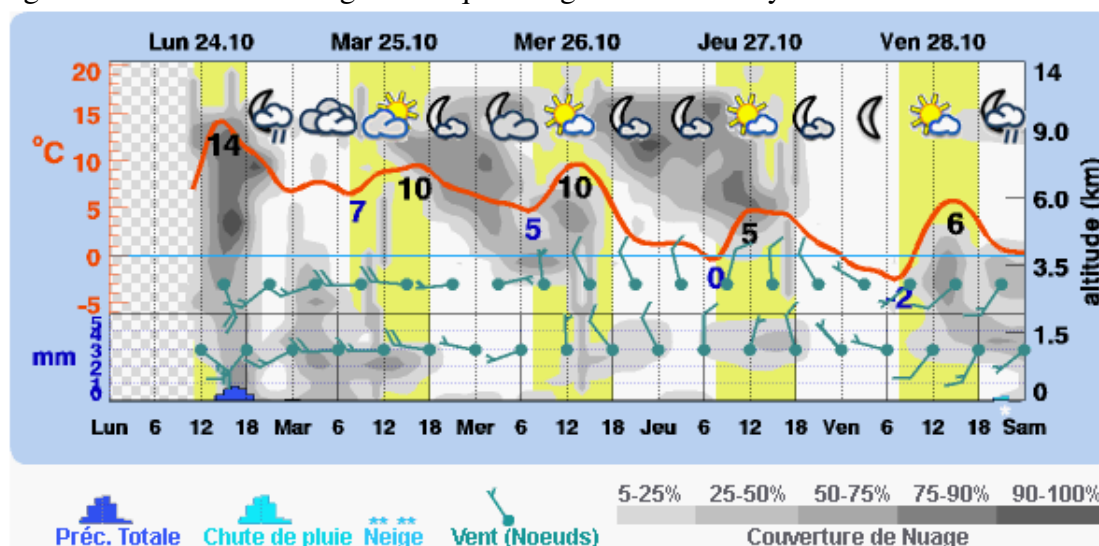


Figure 2.15 – Visualisation générée par Meteoblue indiquant la couverture de nuage (par rapport à l'altitude), la température, la précipitation, la vitesse et la direction du vent. La légende au dessous de la figure indique la signification des symboles utilisés.



2.8.5 Moteur de recherche

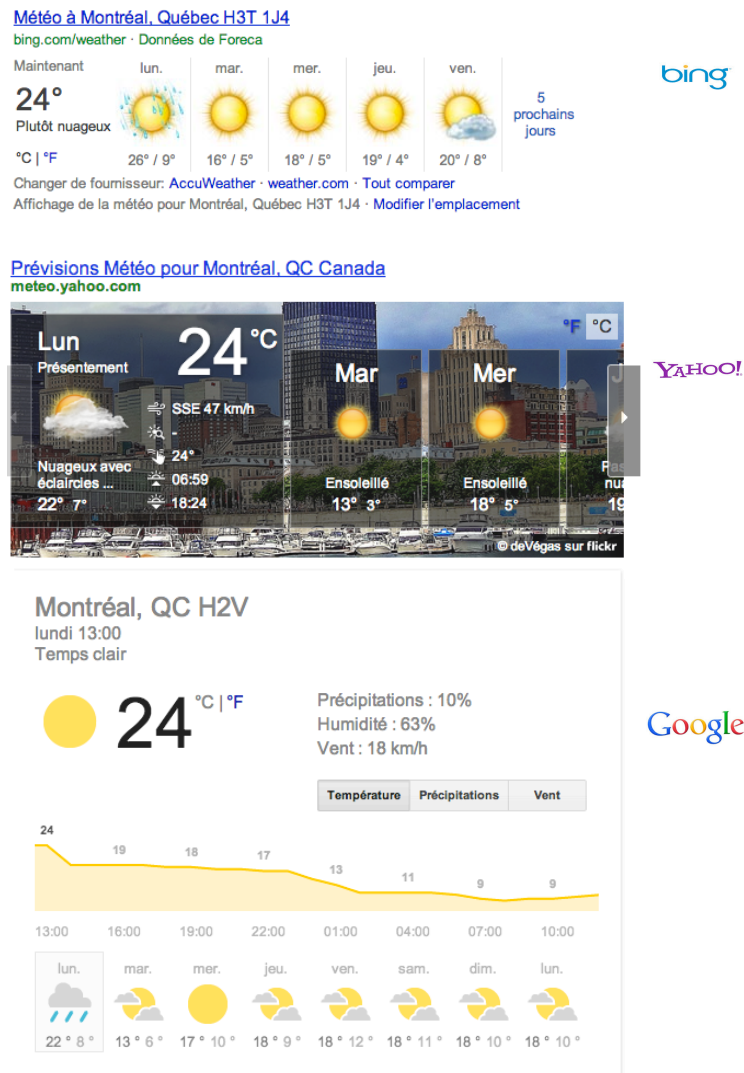
De nos jours, les moteurs de recherches jouent un rôle très important. Ils essaient toujours de faciliter, le plus possible, l'accès à l'information pour les utilisateurs. Une vue d'ensemble des conditions météorologiques est affichée comme premier résultat de la recherche du mot "météo" dans la majorité des moteurs de recherches. Cette vue d'ensemble contient les informations de base de la condition météorologique (température, vent, couverture nuageuse et précipitation). Ces informations sont présentées sous formes iconiques et sous forme de valeurs numériques. La figure 2.16 est un aperçu des résultats affichés par les moteurs de recherche les plus utilisés : bing, google et yahoo. Les moteurs de recherche détectent, à l'aide de la géolocalisation de l'adresse IP, l'emplacement de l'utilisateur et lui affichent directement l'aperçu des conditions météorologiques de son emplacement mais il peut aussi demander « météo Longueuil » pour obtenir la météo d'une autre ville.

Les visualisations générées par les moteurs de recherche sont affichées aux utilisateurs dans un but informatif et ne sont pas exhaustives. Une quantité très minime d'information est communiquée. Aucun texte descriptif n'accompagne ces présentations. Aucune interaction dans le but exploratif n'est offerte. Vers la fin de l'année 2012, le résultat produit par Google se distingue des autres moteurs de recherche. Google génère maintenant une vraie visualisation (figure 2.16). L'utilisateur peut modifier directement la visualisation.

2.8.6 Synopsis

À la section précédente, nous avons étudié des sites présentant de l'information météorologique pour ensuite comparer ces présentations aux méthodes et principes de construction d'une visualisation en général. Nous sommes convaincu que l'interactivité joue un rôle important quand l'information ne doit pas être présentée de la même façon

Figure 2.16 – Résultats affichés par différents moteurs de recherche (bing, google, yahoo et altavista) suite à la requête « météo Montral »



pour tous les utilisateurs. D'autre part, un minimum de principes de base de la construction des visualisations doivent être pris en considération.

Dans les visualisations que nous avons analysées, aucune n'est parfaite. Loin de là, parfois nous nous apercevons que même le minimum n'est pas atteint. Selon notre analyse, nous considérons que les présentations générées par NOAA permettent le mieux aux utilisateurs la perception des informations visualisées. En ce qui concerne des prin-

cipes visant la perception des utilisateurs (principes de Gestalt), NOAA emploie le principe de la similitude pour les catégories d'avertissement et le principe de contour délimiter les États. Le principe du contour est aussi utilisé, dans le même but par EC et Météo France.

Du point de vue interactivité, tous les sites excepté les moteurs de recherche (autre que Google) offrent à leur utilisateur la possibilité de demander plus de détails. EC, Météo France et Meteoblue ne proposent pas d'autre moyen d'interaction. Les visualisations générées par NOAA peuvent être réorganisées en permettant aux utilisateurs d'arranger les données comme leur plait ou de les encoder d'une autre manière.

Concernant les moteurs de recherches, ils affichent un aperçu des conditions météorologiques juste à titre indicatif et ces aperçus ne peuvent être en aucun cas considérés comme étant une visualisation à part entière sauf pour Google qui présente une très bonne visualisation. Dans le tableau 2.I, nous récapitulons les résultats de notre analyse.

2.9 Conclusion

Dans ce chapitre, nous avons décrit les principaux travaux dans le domaine de la visualisation, la perception et l'interactivité. Les visualisations analysées dans ce chapitre ont permis de dégager des règles de conception d'une visualisation dans le domaine de la météo. Dans le prochain chapitre, nous utilisons ces règles et principes pour créer une visualisation qui sera notre point de départ. Nous proposons des méthodes et des approches pour profiler l'utilisateur, dégager ses préférences et générer une visualisation personnalisée qui combine texte et graphique.

Techniques	Principes	EC	NOAA	Météo France	Météo blue	Mot. Rech.
Principes de Gestalt	Proximité					
	Similitude		✓		✓	
	Continuité				✓	✓
	Symétrie					
	Contour	✓	✓	✓		✓
L'interactivité	Sélectionner					✓
	Explorer	✓	✓	✓	✓	✓
	Réorganiser		✓			✓
	Encoder		✓			
	Résumer/Élaborer		✓			
	Filtrer					
	Connecter					
	Annuler/refaire					
	Changer la config.					✓
combinaison texte et graphique		✓		✓		

Tableau 2.I – Tableau récapitulatif des techniques de visualisation utilisés dans les sites web d'EC, NOAA, Météo France, Meteoblue et des moteurs de recherche.

CHAPITRE 3

PERSONNALISATION PAR CLUSTERING

3.1 Introduction

Le défi étudié dans cette thèse est la présentation personnalisée de l'information à l'utilisateur. Nous présentons une approche qui calcule les besoins et les préférences de l'utilisateur selon son profil détecté automatiquement et qui génère un rapport personnalisé. Cette approche s'applique dans les domaines où on génère une grande masse d'informations évoluant dans le temps avec des besoins d'utilisateurs diversifiés. Nous avons appliqué cette approche sur un problème réel dans le domaine de la météo. Environnement Canada (EC) produit une masse énorme d'information météorologique (26Mb deux fois par jour). Cette information est utilisée pour fournir aux Canadiens des renseignements à jour sur les conditions météorologiques. Pour les raisons expliquées dans la section 1.5, nous cherchons à générer des bulletins météorologiques à la demande pour répondre aux besoins spécifiques de l'utilisateur qui le consulte.

3.2 Profil de l'utilisateur

Le profilage automatique de l'utilisateur est courant dans le domaine du e-commerce et de la publicité, afin de lui offrir des produits présentant potentiellement un intérêt. Pour y arriver, plusieurs méthodes (légal) y compris les cookies et l'historique des résultats des moteurs de recherche, sont utilisées.

Connaître le profil de l'utilisateur peut être très rentable pour générer une visualisation appropriée. La recherche sur les profils des utilisateurs et plus précisément la personnalisation de l'information a été largement étudiée dans le domaine de la recherche d'information. Kostadinov (2003) résume les principaux travaux qui ont été faits sur ce

sujet.

Il existe deux types de profil d'utilisateur :

- déterminé par l'utilisateur qui définit ses préférences et ses besoins.
- construit automatiquement en utilisant plusieurs techniques telles que l'historique, le comportement, les règles d'associations, des techniques de classification et des algorithmes de classification.

Mobasher et al. (2002) présente deux techniques de classification de profils d'utilisateur et pages Web consultées, afin de fournir des recommandations personnalisées en temps réel :

- PACT (Profile Aggregations based on Clustering Transactions) rassemble en grappes les transactions similaires de pages consultées par l'utilisateur. Cette technique peut considérer d'autres facteurs pour déterminer le poids des éléments à l'intérieur de chaque profil. Il s'agit notamment de la distance de pages vues par rapport à l'emplacement actuel de l'utilisateur sur le site ou le rang du profil en fonction de son importance.
- ARHP (Association Rule Hypergraph Partitioning), calcule des groupes de pages en fonction du nombre de fois qu'ils apparaissent ensemble dans les *logs* plutôt que de regrouper les transactions elles-mêmes.

La technologie la plus efficace pour profiler un utilisateur est l'utilisation des cookies. Ces derniers ont été mis au point par Netscape, et à l'origine appelés les cookies magiques (MagicCookies sur les ordinateurs Macintosh), mais ils ont été officiellement appelés (Persistent Client State HTTP Cookies) McCarthy et Yates (2010). Parce que le protocole HTTP (connexion entre le client et le serveur) se termine à la fin de la connexion, il est considéré comme « sans état » et une fois que la connexion entre le serveur et le client se termine, rien ne reste dans la mémoire. Les cookies ont été mis au point pour assurer la continuité avec le serveur Web.

À l'origine destinés à servir de panier que l'utilisateur pourrait l'utiliser pour conserver leurs sélections, les cookies ont été détournés pour d'autres fins. En particulier, les cookies tiers ont augmenté les craintes de voir apparaître des collectionneurs de renseignements personnels qui ratissent de grandes bases de données de profils d'utilisateurs Hormozi (2005). Les préoccupations au sujet de la vie privée et l'utilisation de cookies ont été prises en compte lors de la création de cette technologie. Au moment de la mise au point originale de la technologie des cookies, les développeurs Montulli et Netscape étaient conscients des problèmes de confidentialité et ne voulaient pas que les cookies soient utilisés comme un mécanisme de suivi. Afin de s'assurer que les cookies ne seraient pas utilisés comme des mécanismes de suivi, Montulli ne prévoit pas l'attribution d'un numéro d'identification unique qui peut être lu par tous les sites Web. Au lieu de cela, Montulli associe un cookie avec un domaine ou un domaine correspondant, pour s'assurer que les sites Web peuvent interroger ou modifier que leurs propres cookies Shah et Kesan (2009).

Berghel (2002) précise que les cookies ont des propriétés de base, mais qu'ils peuvent stocker pratiquement toute information sur l'utilisateur.

D'autres considèrent le profilage qui peut résulter de l'utilisation des cookies tiers comme la plus grande menace pour la vie privée d'un utilisateur. Un argument à propos de l'utilisation de cookies par des tiers et le profilage est que les cookies tiers créent un profil étendu qui peut ne pas être exact Kristol (2001). Par exemple, si l'utilisateur n'a pas consciemment décidé de visiter le site Web, l'information retournée par leurs cookies tiers ne sera pas précise.

Dans notre projet, notre système de visualisation est destiné à être installé sur le site Web d'Environnement Canada, géré par le gouvernement du Canada pour qui la vie privée¹ des utilisateurs est une priorité, ce qui exclut l'utilisation des cookies.

1. <http://www.ipc.on.ca/English/Privacy-Policy/> : « *The collection of personal information by government organizations must be in accordance with section 38(2) of the Freedom of Infor-*

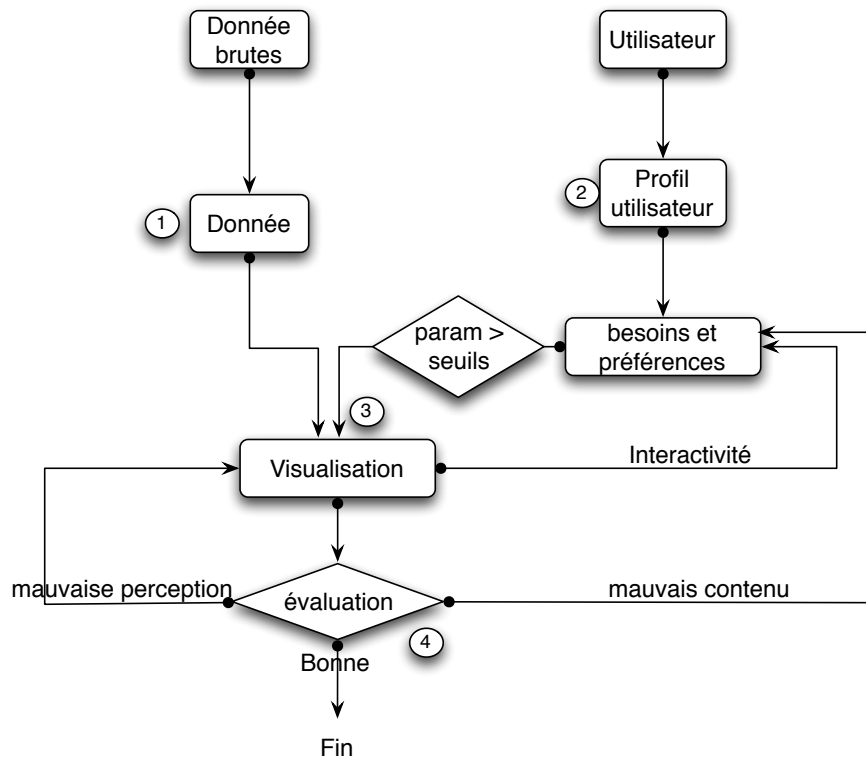


Figure 3.1 – Notre approche prépare les données à partir des données brutes, détermine les besoins et les préférences des utilisateurs, crée la visualisation selon les besoins et les préférences de l'utilisateur et enfin évalue et améliore la visualisation

3.3 Notre proposition

Dans notre application, nous pouvons extraire des informations à partir de deux formats de fichiers (XML et CSV) à partir du site Internet² d'Environnement Canada. Comme indiqué dans l'étape ① de la figure 3.1, nous avons créé un module distinct responsable de la recherche, de nettoyage, de transformation et modélisant les données pour en extraire les informations nécessaires à notre but. Ce module est interrogé par MeteoVis pour obtenir des informations à jour à partir des fichiers produits par EC. Nous prenons en compte le profil des utilisateurs, à savoir leurs besoins et préférences

mation and Protection of Privacy Act (FIPPA). This section of the Act requires government organizations to collect only the personal information that is necessary for the administration of a government program »

2. <http://dd.weatheroffice.ec.gc.ca/meteocode>

afin de générer une visualisation satisfaisante. Cette section correspond à l'étape ② de la figure 3.1. le système doit deviner leurs besoins et leurs préférences fondées sur d'autres informations disponibles.

Nous présentons une méthode, présentée initialement dans Mouine et Lapalme (2012), permettant de personnaliser une visualisation par le clustering des profils des utilisateurs. Les visualisations sont générées dynamiquement par notre système qui enregistre chaque configuration finale choisie par l'utilisateur. L'idée de cette approche est de regrouper les utilisateurs similaires à l'utilisateur étudié dans un même cluster. Le clustering est utilisé pour déterminer les paramètres des utilisateurs. La distance entre l'utilisateur actuel et les autres dans un cluster sera utilisée pour calculer la similarité. Le taux de similarité nous permet d'estimer les besoins et les préférences de l'utilisateur actuel en tenant compte de la similitude du profil de cet utilisateur avec les profils des autres utilisateurs. Le taux de similarité est utilisé pour pondérer les besoins et les préférences. Nous avons fixé un seuil minimum ③ pour chaque paramètre (besoins et préférences). Pour décider quels sont les paramètres qui seront pris en considération dans la visualisation (voir figure 3.1), on ne tient compte que des résultats supérieurs au seuil.

La génération d'une visualisation personnalisée est un mélange d'art et de science. Le côté scientifique s'intéresse principalement à identifier le contenu de la visualisation. D'autre part, le côté artistique est plus intéressé par "comment afficher ce contenu." La perception de l'utilisateur du bon contenu visualisation est l'objectif principal. Comme notre visualisation est personnalisée, le contenu de l'affichage dépend du profil utilisateur. Même si nous avons calculé les besoins et préférences de l'utilisateur, il s'agit d'une prédiction que nous espérons être près des besoins et des préférences réels de l'utilisateur, mais, et sachant la diversité des préférences des utilisateurs, nous proposons deux moyens de prendre en compte le feedback des utilisateurs (l'étape ④).

3.3.1 Critères et paramètres

Pour la génération de la visualisation, nous avons besoin de détecter automatiquement les critères construisant le profil de notre utilisateur pour pouvoir calculer les paramètres qui seront pris en considération dans la visualisation.

3.3.1.1 Critère du profil utilisateur

Les informations que nous avons sur notre utilisateur (sans toucher à sa vie privée) sont celles qu'on peut obtenir en examinant l'information de la connexion HTTP.

- Son emplacement (à l'aide de la géolocalisation de son adresse IP).
- La langue qu'il préfère (français ou anglais) à l'aide des paramètres de son navigateur (variables HTTP USER AGENT et HTTP ACCEPT LANGUAGE).
- L'heure de sa connexion selon son emplacement. Dans le domaine de la météo, la période de la journée pendant laquelle l'utilisateur se connecte peut influencer ses besoins. Un utilisateur qui se consulte la météo le matin a de grandes chances d'être intéressé par les prévisions du jour même (très court terme). Un utilisateur qui consulte la météo le soir a plus de chance d'être intéressé par les prévisions des jours suivants (moyen terme).
- La saison courante (automne, hiver, printemps ou été).

Nous utilisons ces informations pour construire le profil de l'utilisateur. Le tableau 3.I présente les critères que peut avoir un profil.

Tableau 3.I – Critères du profil utilisateur

Information	Critère
Paramètre du navigateur	Langue
Adresse IP	Province et ville
Heure	Période (matin, après-midi et soir)
Date	Saison (hiver, printemps, automne et été)

3.3.1.2 Paramètres de la visualisation

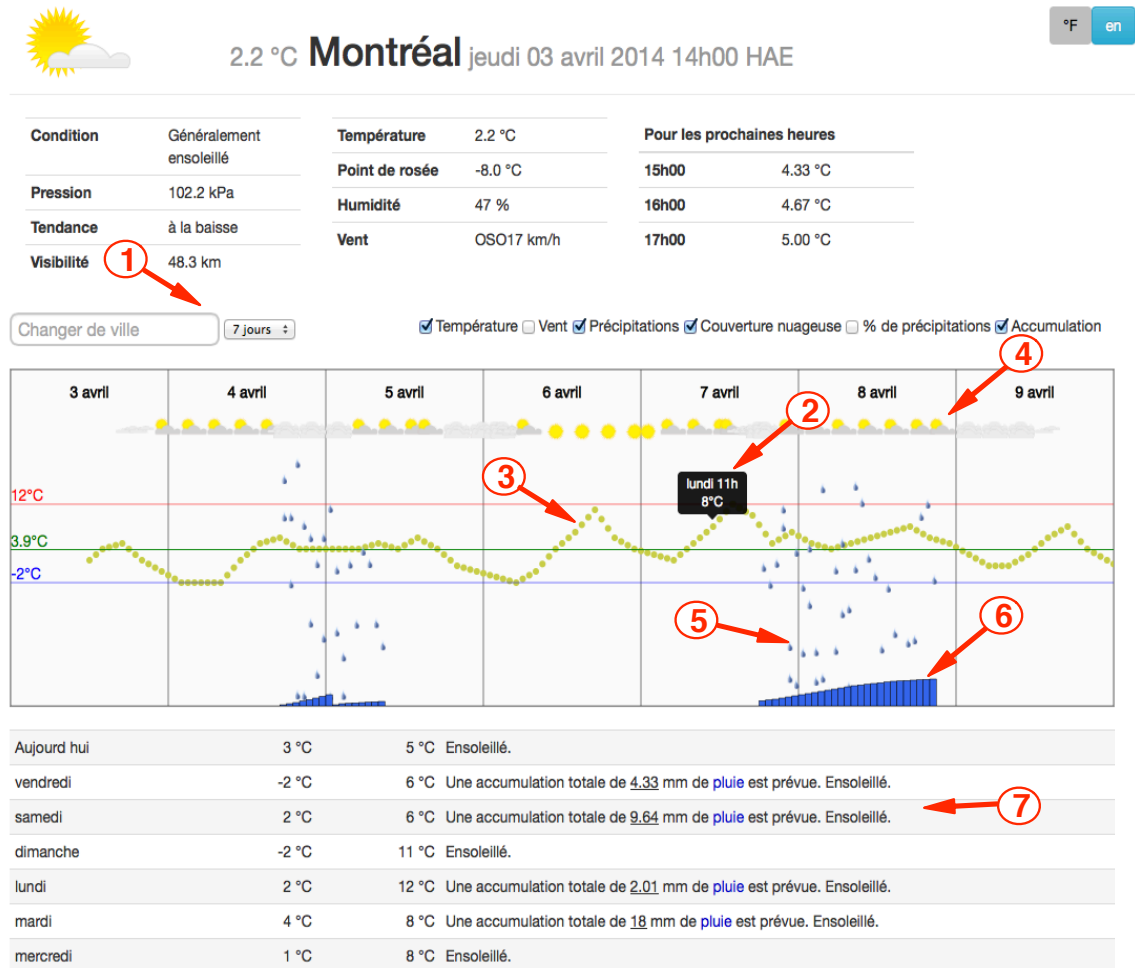


Figure 3.2 – Visualisation montrant certains types d'informations qui peuvent être sélectionnés par l'utilisateur. Les numéros encadrés sont ajoutés ici à titre de référence. Avec le menu des préférences ①, l'utilisateur peut sélectionner une ville, modifier le nombre de jours à afficher et les paramètres à afficher. ② L'utilisateur peut avoir plus de détails sur les aspects en y plaçant le curseur (info-bulle). ③ Température : la façon dont la température est affichée donne une idée de la tendance. Les lignes des maximum et minimum montrent les limites de la température pour la période choisie. ④ La couverture nuageuse. ⑤ type de précipitations : pluie ou neige. La quantité affichée est proportionnelle à la quantité prévue. ⑥ Accumulation : affiche la zone d'accumulation et les précipitations totales (mm) ou de la neige (cm). ⑦ est un texte décrivant l'état de la météo généré automatiquement.

L'information générée par EC (MétéoCode) contient des informations détaillées pour les sept prochains jours pour les paramètres météorologiques suivants : température, vent, type de précipitation, probabilité de précipitation, accumulation. MeteoVis est capable d'afficher toute information contenue dans le MétéoCode pour une période donnée. Nous considérons ces mêmes paramètres météorologiques comme étant ceux de la visualisation auxquels nous ajoutons un paramètre `periode` qui indique le nombre de jours pour lesquels les prévisions seront affichées.

Dans la visualisation de la figure 3.2, un utilisateur peut voir la température ③ : chaque point est la température de l'heure correspondante et indiquant les limites minimales et maximales de la ligne de la température pour la période choisie, la couverture nuageuse ④, le type de précipitations (pluie ou neige) ⑤, (la probabilité de précipitations, l'humidité, le vent ne sont pas affichés sur cette figure) et l'accumulation à un emplacement spécifique ⑥. D'un seul coup d'œil, l'utilisateur perçoit la tendance générale de la température. Les lignes maximum et minimum montrent les limites de la température pour la période choisie. L'utilisateur voit s'il y a de la neige (ou pluie), quand, combien, combien de temps ... Si le vent est affiché (non représenté sur la figure 3.2), la vitesse et la direction du vent sont présentées par la taille et l'orientation d'une flèche.

Dans cette section, nous construisons le profil de l'utilisateur pour qui on génère la visualisation. À partir de ce profil, MeteoVis fixe les paramètres en se basant sur les choix de paramètres des visualisation générées pour des utilisateurs similaires. MeteoVis utilise des techniques d'apprentissage machine (non supervisé) pour regrouper les utilisateurs les plus similaires à l'utilisateur courant.

3.3.2 Clustering

Pour regrouper les utilisateurs selon leur similarité, nous utilisons une méthode statistique d'analyse de données pour partitionner l'ensemble des utilisateurs en grappes

(clusters) d'utilisateurs similaires. Il existe plusieurs algorithmes permettant cela. Parmi les algorithmes nous citons DBSCAN, K-means, EM...

L'algorithme DBSCAN (Density-Based Spatial Clustering of Applications with Noise) regroupe les éléments selon la densité. Les variables que nous devons préciser sont une distance E et un nombre minimum d'éléments pouvant construire un cluster. si ce nombre n'est pas atteint dans un rayon E . Ces éléments sont considérés comme bruit et ne sont attribués à aucun cluster. Le nombre de clusters n'est pas défini à l'avance, mais dépend de la densité des éléments. Le fait qu'un élément risque de n'être attribué à aucun cluster exclut l'utilisation de cet algorithme.

L'algorithme K-means (Hartigan et Wong, 1979) est un des algorithmes d'apprentissage non supervisé pour le problème du clustering. La procédure répartit un ensemble de données parmi un certain nombre K de clusters fixé a priori chacun avec un centroïde. Nous définissons ici K proportionnel au nombre de provinces canadiennes pour plus de précision connaissant la vaste étendue du Canada. La prochaine étape est d'associer chaque utilisateur archivé dans notre base de données (pour qui nous avons déjà généré une visualisation) au centroïde le plus proche. Cet algorithme minimise la fonction objective suivante :

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

où $\|x_i^{(j)} - c_j\|^2$ est une mesure de distance choisie entre un point de données $x_i^{(j)}$ (x présente les utilisateurs dans notre exemple et j sont les critères) et le centre c_j du cluster, c'est un indicateur de la distance des points de données n à partir de leurs centres respectifs des clusters.

Dans ces calculs, nous « éliminons » les différences d'échelle (ordre de grandeur) des variables grâce à une transformation de normalisation. Cette normalisation est la

suivante :

$$V_{norm} = \frac{V_i - \bar{V}}{s}$$

avec :

- $\bar{V}$: la moyenne de l'échantillon.
- s : l'écart type de l'échantillon. C'est la racine carré de la variance.

$$s = \sqrt{\frac{\sum (V_i - \bar{V})^2}{n - 1}}$$

où n est la taille de l'échantillon (nombre d'heures, de saisons...)

Dans notre cas par exemple la variable heure varie de 1 à 24 et la variable saison varie de 1 à 4, les deux vont contribuer de la même manière aux distances à partir desquelles la solution du clustering sera déterminée.

Algorithme

1. Placer les K points dans l'espace représenté par les objets qui sont en cluster.
Ces points représentent les centroïdes des groupes initiaux.
2. Attribuer à chaque objet le groupe qui a le plus proche centroïde.
3. Lorsque tous les objets ont été assignés, recalculer les positions des centroïdes K .
4. Répétez les étapes 2 et 3 jusqu'à ne plus avoir de déplacement de centroïdes.

Cela produit une séparation des utilisateurs en groupes à partir de laquelle la métrique à minimiser peut être calculée.

Le problème du clustering par les K-means est NP-difficile (Dasgupta, 2008) dans le cas général. Pour cela, nous avons décidé d'utiliser cet algorithme avec un corpus de départ (de petite taille) pour fixer les classes, puis faire usage de l'apprentissage supervisé en utilisant la méthode de classement.

3.3.3 Calcul de similarité

À chaque rencontre avec les professionnels d'EC, nous notons qu'ils regroupent principalement les besoins des utilisateurs selon l'emplacement en donnant des exemples tels que les habitants du Nunavut qui s'intéresse principalement au vent (vitesse et direction) et ne s'intéresse pas autant à la température et aux autres paramètres. Nous assignons un poids en conséquent à chaque critère (voir Tableau 3.II). Les informations des profils des utilisateurs et leurs préférences respectives sont sauvegardées (anonymement). Le taux de similarité est donné par :

$$R_j = \sum_{i=1}^n C_i * W_i \text{ avec } j \in [1..m]$$

Où :

- R est le taux de similarité.
- m est le nombre d'utilisateurs
- n est le nombre de critères
- C les critères retenus par l'utilisateur, et W les poids respectifs.

Un exemple de calcul de similarité est montré au Tableau 3.II. Le taux de similarité est utilisé pour pondérer les préférences.

3.3.4 Calcul des préférences

Pour prédire les préférences de l'utilisateur, nous nous appuyons sur les préférences des utilisateurs similaires enregistrées dans notre base de données. Dans la base de données, nous enregistrons le nombre de jours fixé dans la visualisation par l'utilisateur et pour chaque paramètre 1 si le paramètre est utilisé, 0 sinon. Les préférences de chaque utilisateur sont pondérées en fonction du niveau de similitude avec l'utilisateur en cours.

Tableau 3.II – Taux de similarité

Poids	Langue 0.2	Province 0.3	ville 0.3	saison 0.1	période 0.1	% de similarité
utilisateur actuel	fr	Qc	Montréal	hiver	matin	-
utilisateur1	fr	Qc	Québec	automne	après-midi	50%
utilisateur2	fr	Qc	Longueuil	hiver	matin	70%
utilisateur3	en	Qc	Gaspé	automne	matin	40%
utilisateur4	fr	Qc	Montréal	printemps	matin	90%
utilisateur5	fr	Qc	Laval	automne	soir	50%
utilisateur6	fr	Qc	Sutton	été	soir	50%
utilisateur7	en	Qc	Delson	été	matin	40%
utilisateur8	en	Qc	Montréal	printemps	après-midi	60%
utilisateur9	fr	Qc	Gatineau	hiver	matin	70%
utilisateur10	fr	Qc	Montréal	hiver	matin	100%
moyenne						62%

$$P_{wj} = P_j * R_j$$

Où :

- P_{wj} préférences pondéré du critère j .
- R est le taux de similarité.
- P_j Préférence du critère j .

Le tableau 3.III montre un exemple de calcul des préférences. La moyenne de toutes les préférences est comparée à un seuil, si elle est supérieure au seuil, le paramètre en question sera généré dans la visualisation sinon il sera ignoré.

3.4 Génération de la visualisation

Basé sur l'étude des visualisations à partir des sites Web des concurrents d'EC (section 2.8), les principes et les concepts présumés dans ce domaine, en utilisant les principes de perception Ware (2012) et techniques d'interaction Yi et al. (2007), notre

Tableau 3.III – Résultat du calcul des préférences (TOP : type de précipitation et POP : probabilité de précipitation)

	day	température	Vent	TOP	POP	humidité	accumulation
Utilisateur1	2	0.5	0.5	0.5	0.5	0	0
Utilisateur2	3.5	0.7	0	0	0.7	0	0
Utilisateur3	1.2	0.4	0	0	0.4	0	0
Utilisateur4	5.4	0.9	0.9	0.9	0	0.9	0.9
Utilisateur5	2.5	0.5	0.5	0.5	0.5	0	0.5
Utilisateur6	1.5	0	0.5	0.5	0	0	0.5
Utilisateur7	1.6	0.4	0	0.4	0	0	0.4
Utilisateur8	0.6	0.6	0	0	0.6	0.6	0
Utilisateur9	2.8	0.7	0.7	0	0.7	0.7	0
Utilisateur10	6	1	1	1	0	1	0
Total	27.1	5.7	4.1	3.8	3.4	3.2	2.3
Moyenne	4.37	0.57	0.41	0.38	0.34	0.32	0.23
Seuil		0.33					
retenu	4	Oui	Oui	Oui	Oui	Non	Non

système produit un bulletin météorologique³ à l'emplacement initialement déduit de l'adresse IP (voir fig.3.2) avec les besoins et préférences de l'utilisateur en tenant compte du regroupement des utilisateurs similaires.

Contrairement aux rapports, contenant des résumés statiques préparés par EC, cette visualisation contient des informations plus détaillées. Cela est possible parce que nous extrayons dynamiquement les informations précises de 26 Mo de données en fonction du profil de l'utilisateur.

Nous pouvons enrichir la visualisation en ajoutant des informations calculées (moyenne, maximale, minimale ...) afin de simplifier la perception par l'utilisateur.

3. <http://www-etud.iro.umontreal.ca~mouinemo/meteo>

3.5 Evaluation et amélioration de la visualisation

MeteoVis prédit les besoins et les préférences des utilisateurs pour leur générer une visualisation qui répond le mieux à leurs attentes en utilisant le clustering. Néanmoins, on peut améliorer encore le résultat. MeteoVis s’améliore en apprenant du comportement des utilisateurs. MeteoVis est interactif, les utilisateurs ont la possibilité de modifier la visualisation. Nous prenons en compte le résultat de cette interaction pour les profils similaires.

MeteoVis enregistre le couple (profil, préférence) contenant les critères du profil et les paramètres de la visualisation retenue après l’interaction. Ces données seront utilisées dans les prochains calculs avant chaque génération de visualisation. La visualisation générée par MeteoVis est affectée d’une façon directe par le comportement des utilisateurs. MeteoVis utilise l’interaction des utilisateurs avec la visualisation pour prendre en considération la tendance observée sur les profils similaires au profil pour qui on génère la visualisation.

Tableau 3.IV – Résultat de l’interaction avec la visualisation (TOP : type de précipitation et POP : probabilité de précipitation)

	température	vent	TOP	POP	humidité	accumulation
ajouté	-	-	-	-	20%	40%
enlevé	0%	20%	10%	10%	-	-
gardé	100%	80%	90%	90%	80%	60%
moyenne table.3.III	0.57	0.41	0.38	0.34	0.32	0.23
ancien seuil		0.33				
nouveau seuil	0.33	0.35	0.34	0.34	0.31	0.29
retenu (nouveau)	Oui	Oui	Oui	Oui	Oui	Non

3.5.1 Analyse de sensibilité du seuil

L'analyse de sensibilité est utilisée généralement dans les méthodes d'aide à la décision multicritères Brans et al. (1986), Maystre et al. (1994), Roy (1968, 1978). Dans ces méthodes, le décideur fixe des seuils d'une façon arbitraire et ensuite analyse leur robustesse en utilisant l'analyse de sensibilité Maystre et al. (1994), Mena (2001) pour modifier leurs valeurs. Si une faible variation de la valeur du seuil amène un grand changement, alors ce seuil est considéré sensible et la décision non robuste.

Notre approche est inspirée de ce type d'analyse de sensibilité Mouine et Lapalme (2013b). Le seuil détermine si on utilise le paramètre ou non. Nous analysons sa robustesse en utilisant les résultats de l'interaction des utilisateurs avec la visualisation. Nous considérons que :

- si l'utilisateur change l'affichage et ajoute un nouveau paramètre qui n'a pas été utilisé, c'est que cet utilisateur est intéressé par cette information. Le fait que Me-teoVis n'a pas inclus ce paramètre dans la visualisation montre que la valeur du seuil n'a pas permis le choix de ce paramètre. Nous concluons que le seuil était très strict et a conduit à l'exclusion du paramètre en question, sa valeur doit être diminuée ;
- si l'utilisateur modifie la visualisation et la régénère en supprimant un ou plusieurs paramètres, cela signifie que la valeur du seuil pour un paramètre éliminé est très faible et doit être augmenté ;
- si l'utilisateur ne modifie pas les paramètres, la sensibilité des seuils est bonne, la visualisation est robuste et correspond au souhait de l'utilisateur.

Pour évaluer la robustesse⁴ de la visualisation, nous analysons les interactions de l'utilisateur et on calcule les nouveaux seuils selon la formule suivante :

4. La robustesse de la visualisation signifie que le contenu de la visualisation répond exactement aux attentes de l'utilisateur.

Tableau 3.V – Résultat final de la variation des seuils (après 3 itérations)(TOP : type de précipitation et POP : probabilité de précipitation)

	température	vent	TOP	POP	humidité	accumulation
moyenne table.3.III	0.57	0.41	0.38	0.34	0.32	0.23
Seuil final	0.33	0.36	0.34	0.35	0.30	0.30
retenu (nouveau)	oui	oui	oui	oui	oui	non

1. On varie le seuil de ± 0.01 pour chaque 10 % du nombre d'utilisateurs ayant interagi sur ce paramètre.
2. Refaire l'expérience faite dans (Table 3.IV).
3. Refaire l'étape 1 et 2 jusqu'à ne plus avoir plus de 10% d'utilisateurs ayant modifié ce paramètre dans la visualisation.

Table 3.V montre le résultat final de notre exemple. Nous arrêtons la variation des seuils après 3 itérations parce que la visualisation répond aux besoins (préférences) de plus de 90% des utilisateurs.

On peut voir que les seuils ont été révisés à la baisse ou à la hausse selon l'interaction des utilisateurs. Bien que le seuil du paramètre *température* qui satisfait les utilisateurs n'ait pas été modifié, les seuils de *vent*, *TOP* et *POP*) ont augmenté parce que les utilisateurs ne sont pas satisfaits de la visualisation générée pour eux. Le seuil des paramètres *humidité* et *accumulation* était réduit parce que plusieurs utilisateurs voulaient les avoir dans leurs visualisations ce que l'ancien seuil ne permettait pas.

3.6 Conclusion

Le but de notre travail est de proposer des méthodes pour personnaliser la visualisation d'une grande quantité d'informations. Dans notre application, il faut afficher une grande quantité d'informations météorologiques d'une manière simple et s'assurer qu'un usager ait toutes les informations dont il a besoin, et qu'il puisse les analyser. Pour cela,

nous proposons de personnaliser la visualisation pour chaque usager en fonction de son profil que nous devrions détecter automatiquement. Nous calculons les besoins et les préférences de chaque utilisateur en nous basant sur les choix des utilisateurs similaires.

CHAPITRE 4

METEOVIS

L'approche proposée dans cette thèse consiste à personnaliser la génération des bulletins météorologiques pour chaque utilisateur. L'approche que nous proposons se base sur la similitude entre les utilisateurs qui ont déjà utilisé (et interagi avec) MeteoVis et l'utilisateur actuel. Nous utilisons un algorithme d'apprentissage machine non supervisé (clustering) pour regrouper les utilisateurs similaires. Nous utilisons le cluster qui contient cet utilisateur pour calculer ses besoins et préférences. Dans ce chapitre, nous présentons le prototype que nous avons préparé pour réaliser nos expérimentations.

4.1 Architecture

Notre système comprend trois couches (voir figure 4.1). La première correspond aux données générées par EC et hébergées sur le serveur d'EC. La deuxième est hébergée sur le serveur de l'Université de Montréal et contient le module d'extraction de données et une base de données MySQL . La troisième, la couche client est la visualisation générée par MeteoVis .

On extrait les données pour afficher la condition courante à partir de Citypage (voir section 1.3). Cela est dû à la richesse de l'information générée dans citypage concernant la condition courante contrairement aux prévisions qui sont plus détaillées dans le fichier MeteoCode (voir section 1.2). Toutefois, nous utilisons MeteoCode pour extraire les données de la prévision concernant la température, le vent, l'accumulation, le type de précipitation et la probabilité de précipitation.

Le module d'extraction de données est responsable de préparer les données dont le système a besoin. MeteoVis envoie une requête contenant les informations de l'utilisa-

teur pour lequel on veut générer une visualisation et le module d'extraction de données retourne l'identifiant permettant à MeteoVis de retrouver les données dans la base de données MySQL en s'assurant que ces données existent dans la base et à jour. Dans la section suivante nous expliquerons le fonctionnement de ce module.

Un deuxième module responsable du clustering dans la même couche reçoit aussi les informations de l'utilisateur. Ce module se connecte à la base de données des utilisateurs et les regroupe selon leur similarité. Les utilisateurs dans le cluster (groupe) contenant l'utilisateur actuel serviront au calcul de préférences.

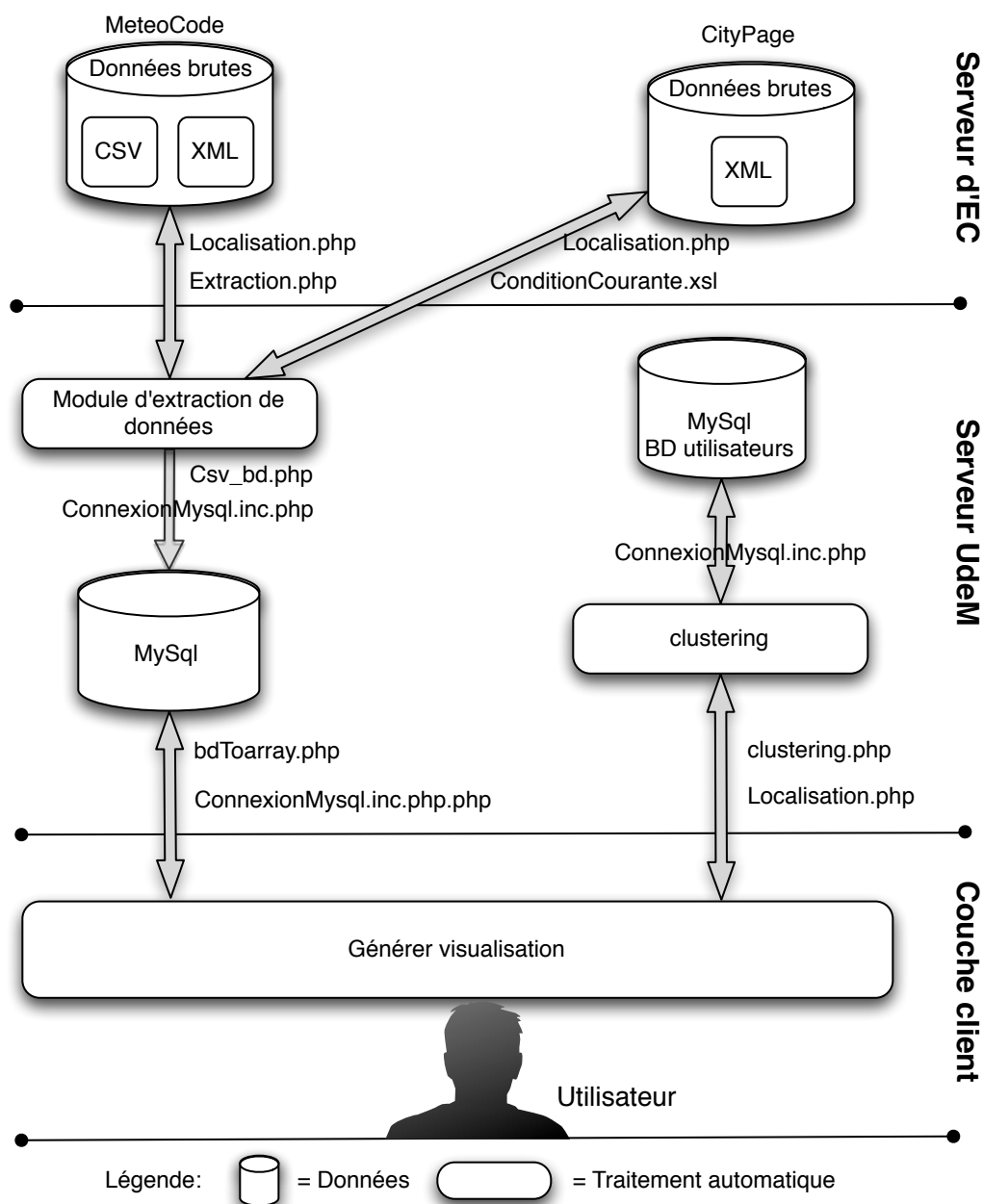
En se basant sur le résultat du clustering et les données retournées par le module d'extraction de données MeteoVis génère une visualisation personnalisée. Cette partie sera discutée dans la section 4.3.3.

4.2 Extraction des données

MeteoVis utilise les données météorologiques générées par EC pour générer une visualisation pour chaque nouvel utilisateur. Afin que les données utilisées soient à jour et correspondent au profil de l'utilisateur, le module d'extraction de données est indépendant.

Pour extraire les données à partir des fichiers CSV et XML (figure 1.2) générés par EC, plusieurs méthodes sont envisageables. La plus simple serait d'extraire à chaque requête de l'utilisateur les données directement à partir des données générées par EC. Cette méthode permet d'alléger le stockage nécessaire au bon fonctionnement de notre système. EC met à jour ses fichiers au moins deux fois par jour. En utilisant cette méthode, nous serions assurés d'utiliser les données météorologiques à jour. Néanmoins, des inconvénients techniques causant un trafic excessif sur les fichiers peuvent alourdir le temps de réponse. Aussi, chaque panne ou non-fonctionnement du serveur d'EC causera un dysfonctionnement de MeteoVis . Nous voulons que MeteoVis soit fiable, rapide et

Figure 4.1 – Architecture du système MeteoVis . Notre système comprend trois couches. La première correspond aux données générées par EC. Les fichiers MeteoCode et City-page hébergés sur le serveur d’EC . La deuxième couche contient le module d’extraction de données et une base de données MySQL hébergée sur le serveur de l’Université de Montréal. La troisième, la couche client est la visualisation générée par MeteoVis .



efficace. À la base, les fichiers contenant les données météorologiques générés et stockés sur le serveur d'EC ne sont pas préparés pour des fins d'utilisation directe. MeteoVis doit être prêt à servir une population de plus de trente millions de personnes ayant l'habitude de consulter fréquemment la météo.

Une deuxième alternative serait de copier toutes les données générées par EC dans une base donnée MySQL. MeteoVis communiquerait directement avec la base de données dans ce cas. Avec cette méthode le module d'extraction de données devrait extraire toutes les informations générées par EC et les injecter dans la base de données MySQL. MeteoVis trouverait alors les informations dont il a besoin directement dans la base de données. Les avantages de cette méthode seraient : un temps de réponse quasi-instantané (entre MeteoVis et la base de données MySQL) ; un risque plus faible de panne causée par le serveur hébergeant les fichiers XML et CSV. Cependant, en utilisant cette méthode nous devrions nous assurer à chaque fois avant d'utiliser les données dans la base qu'elles sont à jour. Comme nous l'avons mentionné précédemment EC met à jour ses données au moins deux fois par jour. À chaque mise à jour nous devons supprimer les données de la base de données et réinjecter les nouvelles données. Une telle opération qui prend en moyenne 18,53 secondes pour extraire l'information à partir de 41 fichiers d'une taille moyenne de 500 Ko chacun. MeteoVis ne peut pas se permettre un tel délai pour générer une visualisation. Le tableau 4.I présente une synthèse des avantages et des inconvénients de chacune des méthodes proposées.

Suite à l'analyse de ces deux méthodes, nous avons décidé de les combiner. Le principe de cette méthode est de faire une copie des données à partir des fichiers d'EC dans la base MySQL à la demande. La base de données (voir figure 4.2) est formée de cinq tables contenant les données météorologiques et une table *Principale* contenant les meta-données et faisant référence aux tables de données météorologiques pour faciliter l'extraction d'information.

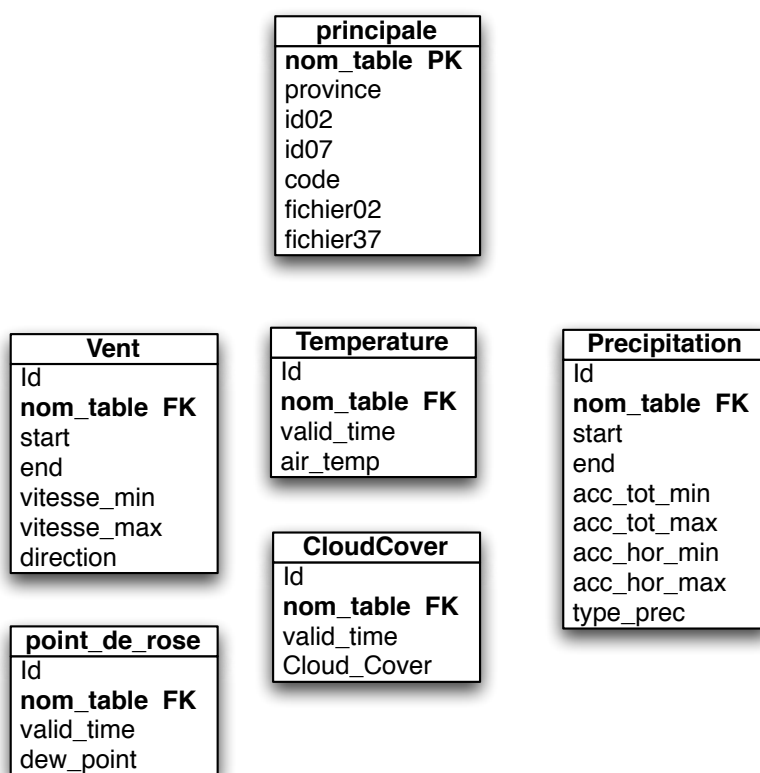
Tableau 4.I – Avantages et inconvénients des deux méthodes d'extraction de données proposées

Méthodes	Avantages	Inconvénients
Première méthode	- Alléger le stockage - Les données sont à jour	- Beaucoup de requêtes. - Une panne sur le serveur d'EC cause l'arrêt de MeteoVis
Deuxième méthode	- Un risque très faible de panne causé par le serveur hébergeant les fichiers XML et CSV. - Une réponse rapide pour toutes les requêtes	- Il faut vérifier à chaque fois que les données sont à jour - À chaque mise à jour on doit re-extraire toutes les données et les injecter dans la base de données (un délai très important : 18,53 secondes)

À chaque fois qu'un nouvel utilisateur se connecte à MeteoVis, ce dernier envoie les informations de localisation de cet utilisateur au module d'extraction de données qui nous renvoie les données météorologiques en appliquant un algorithme (décrit dans l'algorithme 1). Le module d'extraction de données, en se basant sur les informations utilisateurs, s'assure que les données dans la base MySQL concernant cette région sont à jour en comparant la version dans la base de données à la dernière version générée par EC. Si les données pour cette région n'existent pas dans la base de données, le module d'extraction de données récupère, à partir des données générées par EC, la dernière mise à jour pour cette région, les injecte dans la base de données et retourne à MeteoVis le *nom_table* (l'identifiant unique pour cette région dans toutes les tables). Si les données pour une région précise existent dans la base, mais ne sont pas à jour, le module d'extraction de données supprime les données ayant l'identifiant *nom_table* de cette région, récupère les données à jour des fichiers d'EC et les injecte dans la base.

Le module d'extraction de données prend en entrée les données de localisation de l'utilisateur. Ces données sont utilisées pour définir le fichier généré par EC contenant

Figure 4.2 – Shéma de la base de données. Notre base de données se compose de six tables. La table principale contient les informations concernant le fichier MeteoCode utilisé et un champ nom.table qui joue le rôle de clé unique pour retrouver les informations du vent, température, précipitation, point de rosée, couverture nuageuse dans les autres tables.



les données de cette région. Les liens des fichiers générés sont de la forme présentée dans la figure 4.3. Le lien d'un fichier MeteoCode est composé de l'adresse du serveur où stocke EC les données générées de différents types (<http://dd.weatheroffice.ec.gc.ca>) suivi du type de données désiré (MeteoCode). Les prévisions émises pour chaque province se trouvent dans un répertoire spécifique :

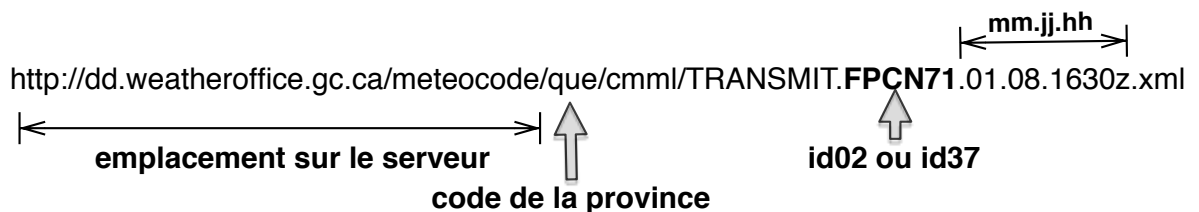
- atl : Nouveau-Brunswick, Nouvelle-Écosse, Île-du-Prince-Édouard, Terre-Neuve-et-Labrador
- ont : Ontario
- pnr : Alberta, Saskatchewan, Manitoba, Territoires du Nord-Ouest, Nunavut

- pyr : Colombie-Britannique, Yukon
- que : Québec

Le CMML est un format XML contenant les prévisions officielles d'EC interpolées toutes les heures. La dernière partie dans l'url est le nom du fichier CMML suivant cette nomenclature : TRANSMIT.FP[A-Z][A-Z][0-9][0-9].MM.DD.HHmmZ[.AMD].xml où :

- MM.DD.HHmm est la date d'émission du bulletin
- Z est une constante qui pour indiquer que l'heure de début de la prévision est en temps UTC.
- AMD : Optionnel. Si présent, cela indique que cette prévision est un amendement à la prévision précédente pour cette région.
- FP[A-Z][A-Z][0-9][0-9] est un code correspondant à la région de prévision

Figure 4.3 – Lien d'un fichier MeteoCode. Ce fichier contient les prévisions émises par le bureau de Québec (répertoire 'que') qui débute le 08 janvier à 1630 UTC ('01.08.1630Z') en format CMML ('.xml')



Le RALI a préparé un fichier XML contenant tous les codes dont nous avons besoin pour chaque région. Un aperçu de ce fichier est présenté dans la figure 4.4.

Une fois le module d'extraction de données a récupéré les informations de la prévision, il les injecte dans la base de données. Notre base de données (figure 4.2) contient une table appelée *principale* qui contient les informations de redirection vers les enregistrements qui nous intéressent. Chaque enregistrement de la table *Principale* contient les informations suivantes :

- la province : La province concernée codée de la même manière que le fait EC (par

Figure 4.4 – Extrait de la liste des codes des fichiers MeteoCode

```

▼<site xmlns:xd="http://www.oxygenxml.com/ns/doc/xsl">
  <city en="Candiac" fr="Candiac" code="s0000046"/>
  <region en="Metro Montréal - Laval" fr="Montréal métropolitain - Laval" code="r71.1"/>
  <station en="Saint-Hubert Airport" fr="Aéroport de Saint-Hubert" lat="+45.50" lng="-73.40"/>
  <province en="Quebec" fr="Québec" code="QC"/>
  <meteocode02 dir="que" id="FPCN71"/>
  <meteocode37 dir="que" id="FPUL55"/>
</site>
▼<site xmlns:xd="http://www.oxygenxml.com/ns/doc/xsl">
  <city en="Calgary" fr="Calgary" code="s0000047"/>
  <region en="City of Calgary" fr="ville de Calgary" code="r15.1"/>
  <station en="Calgary Int'l Airport" fr="Aéroport int. de Calgary" lat="+51.10" lng="-114.02"/>
  <province en="Alberta" fr="Alberta" code="AB"/>
  <meteocode02 dir="pnr" id="FPWG15"/>
  <meteocode37 dir="pnr" id="FPWG53"/>
</site>
▼<site xmlns:xd="http://www.oxygenxml.com/ns/doc/xsl">
  <city en="La Prairie" fr="La Prairie" code="s0000048"/>
  <region en="Metro Montréal - Laval" fr="Montréal métropolitain - Laval" code="r71.1"/>
  <station en="Saint-Hubert Airport" fr="Aéroport de Saint-Hubert" lat="+45.50" lng="-73.40"/>
  <province en="Quebec" fr="Québec" code="QC"/>
  <meteocode02 dir="que" id="FPCN71"/>
  <meteocode37 dir="que" id="FPUL55"/>
</site>
▼<site xmlns:xd="http://www.oxygenxml.com/ns/doc/xsl">
  <city en="Saint-Constant" fr="Saint-Constant" code="s0000049"/>
  <region en="Metro Montréal - Laval" fr="Montréal métropolitain - Laval" code="r71.1"/>
  <station en="Saint-Hubert Airport" fr="Aéroport de Saint-Hubert" lat="+45.50" lng="-73.40"/>
  <province en="Quebec" fr="Québec" code="QC"/>
  <meteocode02 dir="que" id="FPCN71"/>
  <meteocode37 dir="que" id="FPUL55"/>
</site>

```

exemple : pour Québec le code est « que »).

- id02 et id37 : EC génère réellement deux fichiers MeteoCode pour chaque région.

Un fichier MeteoCode02 contenant la prévision des deux prochains jours et un fichier MeteoCode37 contenant les prévisions des 5 jours suivants. EC attribue un identifiant pour chacun d'entre eux. Id02 est l'identifiant du fichier MeteoCode02. (Exemple : FPCN71 : est l'identifiant du fichier MeteoCode02 correspondant à la région de Montréal).

- code de la région (r_RRRR) :
 - r : caractère constant indiquant le début du code de région.
 - RRRR : code alphanumérique indiquant la région de prévision publique. Peut-être d'une longueur arbitraire (ex : 3pe, 10zf, 71.14). (La liste complète peut-

être mise en annexe)

- fichier02 et fichier37 : Deux champs contenant les liens URL des fichiers Meteo-Code02 et MeteoCode37 qui ont servi comme source de données que nous avons dans la base. Ces liens sont utilisés pour vérifier si les informations que nous avons dans la base de données sont extraites de la dernière mise à jour d'EC.
- Nom_table : Clé primaire dans la table principale correspond à une région précise et fait référence à une clé secondaire dans les autres tables de la base de données contenant la prévision météorologique. Nous générons une clé unique de la forme suivante « id02id37code ».

la table *principale* contient toutes les informations nécessaires pour assurer la tâche du mise à jour. L'algorithme 1 décrit les étapes pour s'assurer que les données sont à jour ou les mettre à jour le cas échéant.

```

Fonction MajPrincipale(id02, id37, code, province, fichier02, fichier37) : NomDe-
LaTable
  Enrg : Enregistrement correspondant à l'emplacement de l'utilisateur
  Si (Enrg existe) Alors
    Si (Enrg à jour) Alors
      – result ← Enrg.NomDeLaTable ;
    Sinon
      – Supprimer Enrg et toutes les enregistrements correspondants
        dans les tables descendantes ;
      – Créer Enrg avec les valeurs en paramètre de la fonction ;
      – result ← Enrg.NomDeLaTable ;
    Fin Si
  Sinon
    – Créer Enrg avec les valeurs en paramètre de la fonction ;
    – result ← Enrg.NomDeLaTable ;
  Fin Si
  Retourner result ;
Fin

```

Algorithme 1: L'algorithme de la fonction MajPrincipale

Une fois la base de données à jour, le module d'extraction de données s'arrête et MeteoVis prend la relève. Nous récupérons les informations dans la base de données MySQL en utilisant PHP. Nous faisons une copie dans des tableaux PHP. Ces tableaux servent de point de départ pour la génération de la visualisation.

4.3 Génération de la visualisation

MeteoVis génère une visualisation à la demande pour chaque utilisateur. Un exemple de visualisation est présenté dans la figure 3.2, page 55. La visualisation générée est répartie en trois parties. La première est la condition courante contenant l'état actuel de

la météo. Nous affichons aussi dans cette partie la condition (température et couverture nuageuse) pour les cinq prochaines heures.

La deuxième partie du bulletin climatique est la visualisation de la prévision des prochains jours. Enfin, une description textuelle accompagne le graphique pour mieux décrire la prévision. Les paramètres utilisés dans la visualisation dépendent du profil de l'utilisateur. La personnalisation dans le cas de la génération d'une visualisation dans le domaine de la météo se base sur six paramètres : La période, la température, le vent, le type de précipitation, la couverture nuageuse et l'accumulation. L'utilisateur peut modifier l'affichage initial, y compris la langue et le type de degrés (Fahrenheit/Celsius). L'interaction est immédiate pour tous les paramètres sauf pour le changement de ville. Nous utilisons le framework Bootstrap¹ et la bibliothèque jQuery².

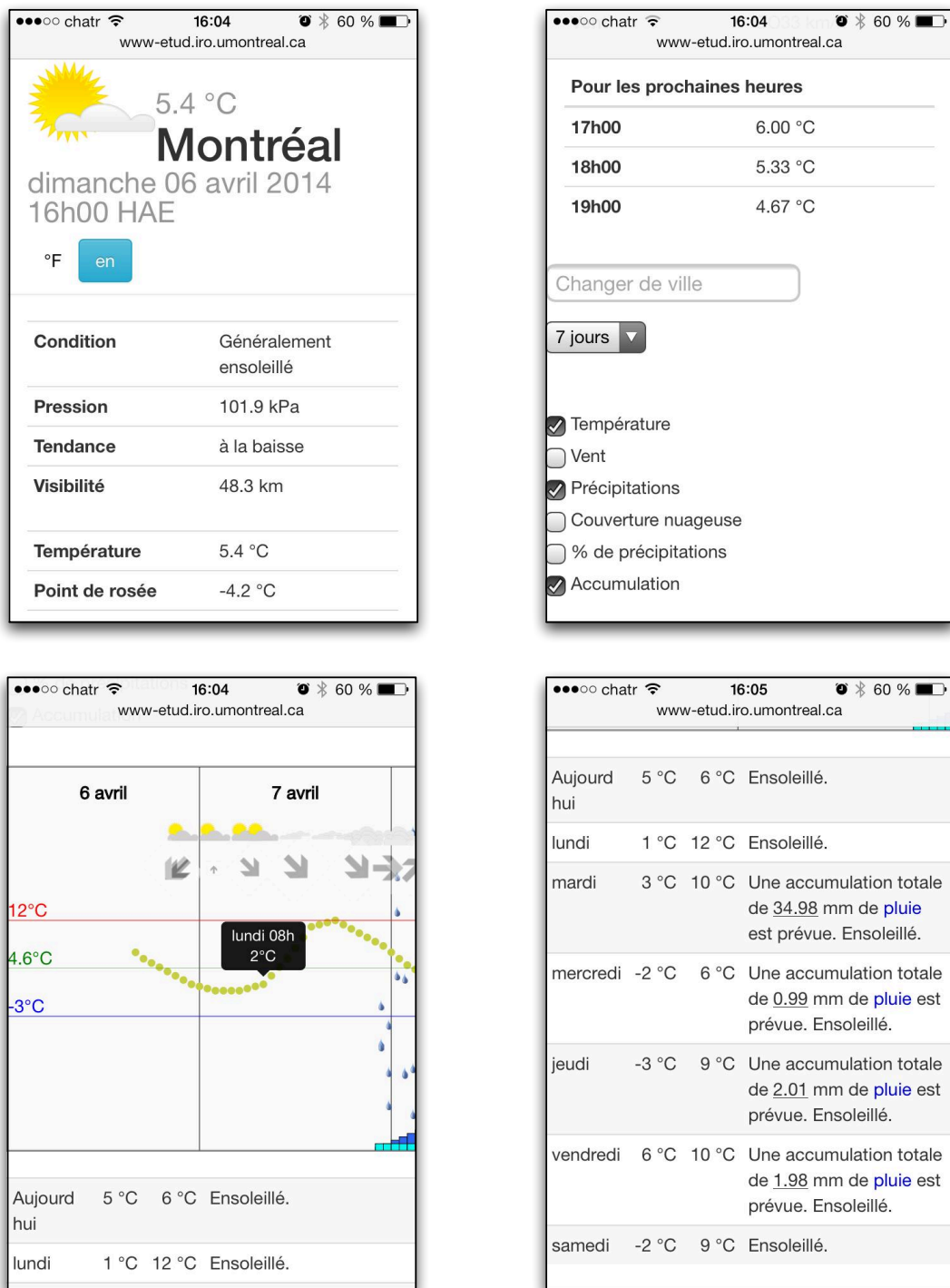
Bootstrap : Bootstrap est un framework CSS qui permet de concevoir des applications web. C'est une collection de codes HTML et CSS. Bootstrap a intégré le soutien jQuery et aussi quelques outils JavaScript populaires. L'utilisation de bootstrap garantit la portabilité et un affichage qui s'ajuste à la largeur de l'écran (y compris les téléphones et les tablettes...). La figure 4.5 est un aperçu de la visualisation générée par MeteoVis sur un téléphone mobile. L'affichage est adapté automatiquement à l'écran d'un téléphone cellulaire. L'utilisateur glisse le doigt verticalement sur l'écran pour parcourir le site. Pour ce qui est du graphique, l'utilisateur peut glisser le doigt horizontalement pour visualiser le reste du graphique.

jQuery : JQuery est une bibliothèque JavaScript gratuite permettant de manipuler aisément la DOM, d'utiliser AJAX, de créer des animations... Son but principal est de simplifier les commandes JavaScript souvent réutilisées.

1. <http://getbootstrap.com>

2. <http://jquery.com>

Figure 4.5 – Aperçu de MeteoVis dans une téléphone mobile. Les vues sont obtenues en glissant sur l'écran



4.3.1 Clustering

Le but de notre travail est de générer une visualisation personnalisée selon le profil de l'utilisateur. L'approche proposée(chapitre 3) se base sur le clustering (section 3.3.2) pour regrouper les profils des utilisateurs similaires à l'utilisateur actuel pour qui on veut générer une visualisation personnalisée.

Une visualisation personnalisée et interactive a été générée pour chacun des utilisateurs pris en compte dans le clustering. Ces utilisateurs ont la possibilité de modifier cette visualisation selon leurs préférences. Nous considérons qu'au moment de quitter MeteoVis les utilisateurs sont « satisfaits ». Basé sur cette hypothèse, le résultat du calcul des préférences (voir section 3.3.4) nous permet de décider des paramètres que nous utiliserons dans la génération de la visualisation pour l'utilisateur actuel.

Le clustering est assuré par le fichier clustering.php. La première étape du clustering charge tous les profils enregistrés dans la base de données. La deuxième étape est d'affecter un poids à chaque critère du profil utilisateur. La troisième étape est de regrouper les utilisateurs dans des clusters. En tenant compte seulement des utilisateurs dans le même cluster que notre utilisateur, MeteoVis calcule les préférences de l'utilisateur pour qui on veut générer la visualisation.

4.3.2 Condition courante

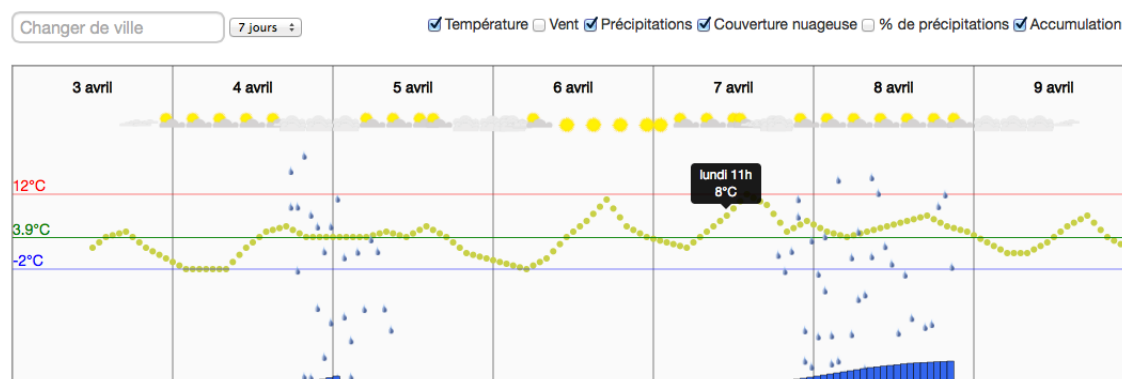
Pour générer la condition courante (voir figure 4.6), nous utilisons le fichier City-Page contenant des informations de la météo actuelle que nous ne pouvons pas trouver dans MeteoCode comme la pression et la visibilité. Nous avons préparé une feuille de style XSLT pour extraire les informations du fichier Citypage . Nous avons enrichi la présentation de la météo actuelle par une partie de la condition des prochaines heures. On récupère directement la condition des prochaines heures de la base de données. La température et la couverture nuageuse de l'heure actuelle et des prochaines heures.

Figure 4.6 – Zoom sur la figure 3.2 présentant la condition courante générée par Meteo-Vis



4.3.3 Graphique

Figure 4.7 – Le graphique généré par MeteoVis (cette figure présente une zoom sur la partie centrale de la figure 3.2)



Cette partie est placée au centre de la page et considérée comme la plus importante. La visualisation est constituée de 3 parties, mais notre travail insiste sur cette partie graphique.

Nous utilisons le résultat du clustering dans cette partie pour définir le contenu du graphique. Si le calcul montre que les utilisateurs similaires préfèrent ne pas afficher le vent par exemple, nous n'afficherons pas ce paramètre pour l'utilisateur actuel (dans le tableau 3.III nous avons donné un exemple de résultat final du choix de paramètre à

afficher) .

Nous avons vu au chapitre 2 plusieurs techniques, méthodes et approches de visualisation. Nous avons aussi analysé plusieurs sites web offrant le service de prévision météorologique, dont le site d'EC. Le tableau 2.I présente une synthèse de cette analyse. Nous créons notre visualisation sur une base scientifique en suivant tous les points de cette analyse pour que la visualisation soit la plus complète. Nous prenons en considération le point de vue utilisateur en essayant de maximiser sa perception pour la visualisation. Nous mettons en évidence (pour un traitement préattentif) les informations importantes.

MeteoVis génère le graphique (voir figure 4.7) en utilisant le formalisme SVG (Scalable Vector Graphics).

SVG La spécification SVG est un standard ouvert élaboré par le World Wide Web Consortium (W3C) depuis 1999. Les images SVG et leurs comportements sont définis dans des fichiers XML. Cela signifie qu'ils peuvent être recherchés, indexés, scénarisés et, si nécessaire, comprimés. Comme ce sont des fichiers XML, des images SVG peuvent être créées et éditées avec n'importe quel éditeur de texte.

Tous les grands navigateurs Web ont au moins un certain degré de soutien et d'interprétation du balisage SVG directement, y compris Mozilla Firefox, Internet Explorer 9, Google Chrome, Opera et Safari. Toutefois, les versions antérieures de Microsoft Internet Explorer (IE) ne prennent pas en charge nativement SVG.

Les applications Web utilisant SVG permettent aux utilisateurs d'entrer leurs propres données, modifier des données, ou même produire de nouveaux graphiques. Puisque les données sont résidentes sur l'ordinateur client, l'interactivité est presque instantanée.

Le code source des images SVG est défini dans un fichier XML.

Tous les fichiers de données générés par EC sont aussi en XML.

Le graphique généré dépend du résultat du clustering et utilise les informations sur la prévision enregistrées dans la base de données par le module d'extraction de données. La décision d'afficher ou non un paramètre dans le graphique est prise par MeteoVis en conséquence du calcul des besoins et préférences de l'utilisateur.

Toutefois, si le graphique généré ne répond pas « exactement » à ce à quoi l'utilisateur s'attendait, il a toujours la possibilité d'interagir avec la visualisation en modifiant la période (de 1 à 7 jours) et cocher (ou décocher) les cases correspondant aux paramètres de la visualisation (température, vent, type de précipitation, couverture nuageuse, pourcentage de précipitation, accumulation).

MeteoVis donne la possibilité à l'utilisateur d'aller chercher plus de détails concernant les informations générées.

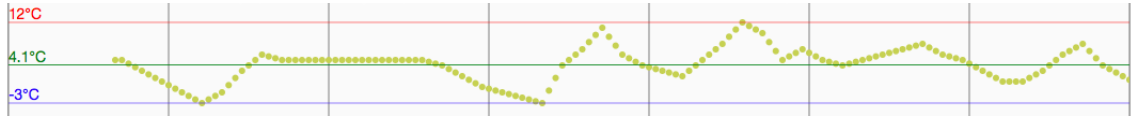
Figure 4.8 – Le vent



Le vent est présenté par des flèches indiquant la direction. La taille de la flèche donne une idée à l'utilisateur de la vitesse du vent. Plus la flèche est grande plus la vitesse du vent est forte. En plaçant le curseur sur une des flèches, MeteoVis affiche une infobulle contenant l'heure de début et de fin du vent à cette même vitesse, la vitesse minimum et maximum et la direction. Nous utilisons la même flèche, mais nous varions sa taille et sa direction en fonction de la vitesse et de la direction d'où l'avantage d'utiliser SVG qui permet des rotations, translations et changement d'échelle. Selon le principe de similitude de Gestalt, l'utilisateur reconnaît que toutes les flèches indiquent un même paramètre à valeurs différentes.

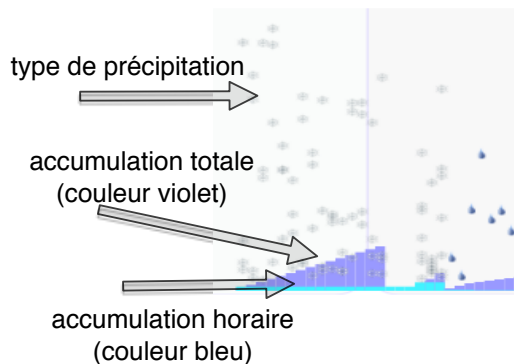
La température est représentée par un point pour chaque heure (figure 4.9). Selon le principe de continuité (les principes de Gestalt), nous sommes plus enclins à construire des entités visuelles à partir d'éléments lisses et continus. Cela s'ap-

Figure 4.9 – La température



plique à l'ensemble des points indiquant la température et constituant une courbe donnant une idée générale de la tendance de la température pour la période choisie. En positionnant le curseur sur l'un des points température MeteoVis affiche un complément d'information dans une infobulle pour montrer à l'utilisateur le jour, l'heure et la valeur de la température pour ce point. Afin d'améliorer la perception de l'utilisateur, nous incluons dans le graphique de l'information déduite des données de prévision générées par EC. MeteoVis affiche deux lignes pour montrer respectivement le minimum et le maximum que la température atteindra pour la période sélectionnée et une troisième ligne indiquant la moyenne. En positionnant le curseur sur une de ces lignes l'utilisateur recevra une infobulle contenant : « *La température [maximum / minimum] prévue pour cette période est :[valeur]. Cette température est prévue pour le [jour] à [heure]* »

Figure 4.10 – Présentation de l'accumulation et du type de la précipitation



L'accumulation est présentée par deux barres superposées pour chaque heure. Ces deux barres sont de couleurs différentes (bleu et mauve) et représentent respectivement

l'accumulation horaire et totale. L'utilisateur pourra facilement distinguer l'accumulation totale et l'accumulation horaire en se basant sur le principe de contour. Les informations optionnelles affichées dans l'infobulle sont l'heure, la valeur de l'accumulation horaire et totale et le type de précipitation.

Le type de précipitation est une animation de flocons de neige ou de la pluie descendant du niveau de la couverture nuageuse au niveau de l'accumulation facilitant la compréhension des autres composantes de la visualisation et donnant l'impression que la précipitation (pluie ou neige) descend des nuages et s'accumule par terre.

La couverture nuageuse est présentée par des icônes sur un laps de temps de trois heures. Les informations numériques dans la base de données concernant la couverture nuageuse est un nombre entier variant de 0 à 9. Nous ne pouvons pas afficher une image pour chaque heure (manque d'espace). Nous affichons une image pour chaque trois heures dans le même intervalle de couverture nuageuse (voir tableau 4.II).

Tableau 4.II – Les intervalles pour la couverture nuageuse

nombres	Image
0	
1 et 2	
3 et 4	
5 et 6	
7, 8 et 9	

4.3.4 Génération de texte

Au départ, nous avons planifié l'utilisation de SimpleNLG-ENFR pour cette partie. SimpleNLG-ENFR est une bibliothèque Java permettant de générer automatiquement du

Figure 4.11 – Le texte généré par MeteoVis

Aujourd'hui	3 °C	5 °C Ensoleillé.
vendredi	-2 °C	6 °C Une accumulation totale de <u>4.33</u> mm de pluie est prévue. Ensoleillé.
samedi	2 °C	6 °C Une accumulation totale de <u>9.64</u> mm de pluie est prévue. Ensoleillé.
dimanche	-2 °C	11 °C Ensoleillé.
lundi	2 °C	12 °C Une accumulation totale de <u>2.01</u> mm de pluie est prévue. Ensoleillé.
mardi	4 °C	8 °C Une accumulation totale de <u>18</u> mm de pluie est prévue. Ensoleillé.
mercredi	1 °C	8 °C Ensoleillé.

texte en anglais et en français. L'intégration de cette bibliothèque dans notre prototype nécessite plus de temps que prévu. Pour cela, nous avons opté à l'utilisation des gabarits de phrases. Nous avons préparé un module de génération de texte basé sur des « gabarits de phrases ». Ce module génère des portions de phrases préparées selon l'entrée (voir figure 4.11). L'expression générée décrivant l'état de la météo est divisée en plusieurs parties, chaque partie désignant un paramètre de la visualisation. Le choix de la partie utilisée dépend de la valeur numérique de l'entrée. Les paramètres retenus pour la visualisation (résultat du clustering) jouent un rôle important dans la détermination du contenu pour la génération de texte. Le texte ne décrit que les paramètres retenus par le système. Le générateur de texte ignore aussi les paramètres dont les valeurs sont jugées de faible importance (exemple : vitesse du vent = 5 km/h). De même, les paramètres qui ne sont pas retenus par le système peuvent être utilisés si leurs valeurs sont jugées de haute importance et seront affichés sous forme d'avertissement (exemple : valeur totale prévue d'une averse de neige > 25 cm). Pour construire la phrase, le système prépare des propositions pour chaque paramètre retenu selon les critères décrits précédemment et les affiche dans la zone de description du jour correspondant.

EC utilise le même concept pour générer la description textuelle dans la version du site actuel. Comme nous l'avons vu dans le chapitre 1, EC prépare à l'avance les bulletins climatiques. Plus exactement, prépare un fichiers Citypage pour chaque région (854

Tableau 4.III – Exemple d'apprentissage des gabarits de phrases utilisés par EC

temp (max)	min	Vent	POP (%)	TOP	Texte
-6	-8		60	snow	Nuageux avec 60 pour cent de probabilité d'averses de neige. Minimum moins 8. Maximum moins 6.
-3	-7		40	snow	Nuageux avec 40 pour cent de probabilité d'averses de neige. Minimum moins 7. Maximum moins 3.
-5	-9			snow	Ensoleillé. Minimum moins 9. Maximum moins 5.
-1	-5		40	snow	Nuageux avec 40 pour cent de probabilité d'averses de neige. Minimum moins 5. Maximum moins 1.
4				rain	Pluie intermittente. Maximum plus 4.
	2	20-40 15-0		rain	Nuageux. Pluie intermittente débutant en soirée. Vents du sud-ouest de 20 km/h avec rafales à 40 devenant légers en soirée. Températures stables près de plus 2
	3	15-0 20-40		rain	Alternance de soleil et de nuages. Devenant nuageux le matin. Vents devenant du sud-ouest à 20 km/h avec rafales à 40 en mi-journée. Maximum plus 3.

régions). Ce fichier contient les données de la condition courante, les données (abrégés) de la prévision des sept prochains jours, et une description textuelle (voir section 1.3). Pour préparer les gabarits de phrases nous nous sommes basés sur les phrases générées par EC. Nous avons préparé un script qui, à partir de Citypage, reliant les données numériques aux phrases générées. Le tableau 4.III montre un exemple du résultat du

script. En suivant cette méthode nous avons pu constituer une bibliothèque de gabarits et nous avons pu dégager les règles utilisées pour le choix des gabarits. Cette manipulation a été réalisée aussi pour les bulletins d'EC qui sont générés en anglais. Ainsi on garantit que les textes français et anglais sont les mêmes que ceux préparés déjà par EC pour ses prévisions habituelles.

4.4 Conclusion

MeteoVis permet aux utilisateurs d'interagir avec la visualisation en permettant de rajouté des paramètres qui n'ont pas été affichés suite à la phase du clustering, modifier la période affichée, demander plus de détails ou bien enlevé des parties qu'il juge non consistantes. Ces modifications sont enregistrées dans la base de connaissance de MeteoVis accompagnées des informations du profil de l'utilisateur. Cette même base de données est utilisée pour le calcul des besoins et préférences des nouveaux utilisateurs. Ce qui influe directement sur le prochain calcul pour améliorer le contenu généré.

CHAPITRE 5

ÉVALUATION ET AMÉLIORATION

Les graphiques sont utiles pour résumer et communiquer des informations numériques trouvées dans les bulletins météo. La tâche de générer des graphiques ne peut pas être réduite à l'encodage d'une masse d'informations, on doit prendre en compte le décodage à faire par l'utilisateur. Si l'utilisateur ne peut pas décoder les informations, on considère que la génération graphique a échoué. Il convient de prendre en compte la perception humaine. Robbins (2012) explique comment créer de meilleures visualisations en tenant compte de tous les paramètres (choix de type de graphique, la quantité d'informations, le choix des attributs de style...) et de la perception envers le graphique.

Dans notre analyse de l'existant effectuée au chapitre 2 (section 2.8), nous nous sommes positionnés du point de vue utilisateur pour évaluer les visualisations existantes. Nous nous sommes basés sur les techniques et principes étudiés dans le même chapitre en prenant en considération l'étude faite par (Chen, 2005) sur les dix problèmes basés sur l'utilisateur, en nous attardant sur les trois premiers : la facilité d'utilisation, la perception et les connaissances préalables de l'utilisateur.

Ces trois problèmes étant centrés sur l'utilisateur, connaître le profil de l'utilisateur peut être très rentable pour générer une visualisation appropriée.

MeteoVis tente de prédire les préférences des utilisateurs. Les rapports générés pour chacun sont interactifs. Il est très important que les utilisateurs puissent modifier la visualisation en fonction de leurs goûts et besoins.

L'approche que nous avons proposée dans cette thèse nous permet de générer une visualisation personnalisée pour un utilisateur sur lequel nous avons peu d'information sur son profil en nous basant sur les préférences des utilisateurs enregistrés dans la base de données avec un profil similaire à celui de l'utilisateur pour qui on veut générer la

visualisation.

Dans ce chapitre, nous évaluons la visualisation générée par MeteoVis par rapport aux techniques d'interactivité et leurs impacts sur la perception. Nous revenons sur chaque technique pour réaliser une analyse similaire à celle réalisée dans la section 2.8. Nous évaluons par la suite l'intégration des principes de Gestalt dans la visualisation. Une grande partie de ce chapitre est consacrée à une étude de l'utilisation de MeteoVis par des usagers.

5.1 Interactivité

MeteoVis génère une visualisation personnalisée selon les besoins et préférences de l'utilisateur. Néanmoins, nous lui donnons la possibilité de changer cette visualisation en l'adaptant à ses besoins et préférences.

Pour que la visualisation de données qui varient dans le temps soit pratique, l'usager doit se concentrer sur une plage de temps et sur une partie des données. En conséquence, les usagers doivent pouvoir consulter de manière interactive les différentes parties du domaine du temps et inclure des variables alternatives dans le codage visuel pour arriver à une vue globale des données.

Le premier but de l'interactivité est de mieux répondre aux besoins de l'utilisateur et de rendre la visualisation plus personnalisée. Nous utilisons aussi l'interactivité pour mieux prédire les préférences des utilisateurs dans le futur. À la fin de la visite de l'utilisateur à notre site web, nous enregistrons dans une base de données MySQL les informations du profil et le résultat de l'interaction.

Dans l'analyse des visualisations existantes (section 2.8), nous avons évalué l'utilisation des techniques d'interactivité pour les sites web étudiés (tableau 2.I). Nous montrons comment nous avons utilisé dans MeteoVis la majorité des techniques d'interactivité étudiées à la section 2.4 et rappelées dans le tableau 5.1 :

L'interactivité						
Principes	EC	NOAA	Météo France	Météo blue	Mot. Rech.	MeteoVis
Sélectionner					✓	✓
Explorer	✓	✓	✓	✓	✓	✓
Réorganiser		✓			✓	✓
Encoder		✓				✓
Résumer/Élaborer		✓				✓
Filtrer						
Connecter						
Annuler/refaire						✓
Changer la config.					✓	✓

Tableau 5.I – Tableau récapitulatif des techniques d'interactivité utilisés dans les sites web d'EC, NOAA, Météo France, Meteoblue et des moteurs de recherche et MeteoVis .

Sélectionner : Le système se base sur l'adresse IP de l'utilisateur pour générer un bulletin météorologique pour l'emplacement estimé. MeteoVis permet à l'utilisateur de « sélectionner » un autre emplacement.

Explorer : Le menu ① dans la figure 3.2 permet à l'utilisateur de choisir le nombre de jours à « explorer » dans la visualisation. MeteoVis calcule le nombre de jours selon les choix des utilisateurs dans le même cluster. Le nombre de jours affiché dans la visualisation est la partie entière supérieure de la moyenne des nombres de jours choisis par les utilisateurs enregistrés dans la base de données et appartenant au même cluster. La formule utilisée est la suivante :

$$NbrJour = \lceil \frac{\sum_{i=1}^n Nbr\ jours User_i}{n} \rceil$$

Avec :

- $Nbr\ jours User_i$ est le nombre de jours choisi par le i ème utilisateur dans le cluster.
- n le nombre d'utilisateurs dans le cluster.

Réorganiser : MeteoVis affiche beaucoup d'information évoluant dans le temps. La personnalisation de la visualisation, sujet principal de notre recherche, permet de décider quelles informations seront affichées selon le profil de l'utilisateur. Néanmoins, l'utilisateur a toujours la possibilité de choisir un arrangement différent en modifiant les paramètres de la visualisation.

Changer la configuration : L'utilisateur peut changer la configuration de la visualisation et agir sur tous les paramètres. En utilisant le clustering et le calcul des préférences, MeteoVis génère une visualisation contenant les paramètres résultant de ce calcul.

Résumer/ Élaborer : Lors de l'analyse visuelle, les usagers ont besoin de regarder certaines choses en détail, tandis que pour d'autres des représentations schématiques sont suffisantes. Nous appliquons le principe d'interactivité « Résumer / Élaborer » sur la plupart des représentations graphiques de la visualisation. L'utilisateur a la possibilité d'élaborer le niveau de granularité des données présentées graphiquement en positionnant le curseur sur la partie demandée. Une infobulle apparaît pour fournir plus de détails. La figure 5.1 montre un exemple d'infobulle.

Annuler/Refaire : Nous avons indiqué que l'utilisateur peut changer la configuration de la visualisation comme bon lui semble. Il est aussi important que ce dernier puisse « annuler/refaire » toutes les manipulations pour avoir un résultat qui répond à ses besoins et qui lui permet d'analyser et de percevoir les informations contenues dans le graphique.



Figure 5.1 – Exemple d’infobulle.

5.2 Principes de Gestalt

Le but d’une visualisation est d’afficher une grande masse d’informations à l’utilisateur d’une manière claire et compréhensible. Les principes de Gestalt ¹ (section 2.3.2) constituent des « règles » que le créateur d’une visualisation peut utiliser pour faciliter la perception de la visualisation par l’utilisateur. Dans l’évaluation de notre bulletin météorologique, nous analysons l’utilisation de chaque principe et rappelé au tableau

5.II

Principes de Gestalt						
Principes	EC	NOAA	Météo France	Météo blue	Mot. Rech.	MeteoVis
Proximité						✓
Similitude		✓		✓		✓
Continuité				✓	✓	✓
Symétrie						✓
Contour	✓	✓	✓		✓	✓

Tableau 5.II – Tableau récapitulatif des principes de Gestalt utilisés dans les sites web d’EC, NOAA, Météo France, Meteoblue et des moteurs de recherche et MeteoVis .

1. Proximité, similitude, continuité, symétrie, contour

La proximité est utilisée pour regrouper les flocons de neiges et les gouttes de pluie (⑤ dans la figure 3.2) en transition verticale dans visualisation pour définir un ensemble visuellement regroupé et considéré comme un seul élément visuel ce qui diminue la distraction visuelle.

La similitude : nous présentons quatre types d'accumulation (⑥ dans la figure 3.2) dans la visualisation (accumulation horaire, accumulation totale, accumulation de pluie et accumulation de neige). La présentation de ces quatre types est similaire à différents endroits (selon la date et l'heure de la précipitation). Nous présentons l'accumulation sous forme d'une barre pour chaque heure. Les barres indiquant l'accumulation horaire sont de couleur bleue. Chaque barre d'accumulation horaire est superposée sur une barre d'accumulation totale de couleur violet. Selon la loi de similitude, les éléments sont regroupés perceptuellement s'ils sont semblables, d'où l'utilisateur perçoit à la première étape du traitement cognitif (appelé aussi traitement pré attentif) que les blocs de barres représentent l'accumulation. Et dans la deuxième étape du traitement cognitif, l'attention visuelle permet à l'utilisateur d'examiner les éléments plus en détail.

La continuité : chaque point dans ③ de la figure 3.2 représente la température de l'heure correspondante. Sachant que la température ne peut pas subir une variation énorme d'une heure à l'autre, l'ensemble des points assure une continuité dans l'espace ce qui constitue une entité visuelle. À l'étape du traitement préattentif l'utilisateur aperçoit une courbe indiquant la tendance de la température sur la période affichée. Le traitement séquentiel dirigé par les buts étant la deuxième étape du processus de perception permet à l'utilisateur, s'il en a besoin, d'aller exploiter la température pour une heure ou période précise.

La symétrie : Chaque jour de la période affichée est présenté d'une façon symétrique

(verticalement) au prochain jour. La loi de symétrie adopte l'idée qu'en voyant deux éléments sans lien qui sont symétriques, inconsciemment nous les intégrons dans un objet percept. Nous avons bénéficié de cette loi de symétrie pour faciliter la perception de l'utilisateur pour la visualisation pendant les deux étapes du traitement de l'information et de la perception visuelle humaine. À l'étape du traitement préattentif l'utilisateur perçoit un bloc représentant la période en totalité et par conséquent, il a une idée générale de l'état de la météo pour cette période. Automatiquement, la concentration visuelle est perceptive de l'utilisateur se dirige par la suite au jour dont il a besoin. La différence des couleurs et la séparation en plusieurs éléments symétriques aideront l'utilisateur à pointer son champ visuel sur la journée dont il a besoin.

Le contour : MeteoVis trace une ligne horizontale de couleur rouge au niveau de la température maximum pour la période affichée et une ligne de couleur bleue au niveau de la température minimum. Ces deux lignes délimitent visuellement les points représentant la température dans la visualisation. Selon la loi du contour, partout où un contour est vu, il y a tendance à diviser les régions de l'espace en « intérieur » et « extérieur » du contour. Sachant que nous indiquons sur chaque ligne la valeur numérique de la température correspondante, l'utilisateur perçoit directement la température maximum et minimum pour cette période.

5.3 Étude empirique : Rétroaction des utilisateurs

Nous voulons dans cette section mesurer la performance de notre visualisation. Kosara et al. (2003) souligne que peu importe l'efficacité d'une technique de visualisation ou la motivation de la théorie, si elle ne donne pas d'informations de manière efficace, elle n'est pas d'une grande utilité. Quatre des plus grands chercheurs dans le domaine de la visualisation ont donné leurs avis dans Grinstein et al. (2003) sur la question : « Which

comes first, usability or utility ? ». Les quatre ont souligné l'importance de l'une autant que l'autre. L'un de nos rôles en tant que chercheur en visualisation de l'information est de fournir des preuves convaincantes de l'utilité et la convivialité, ce qui est difficile pour toute nouvelle technologie, mais présente des défis particuliers pour la visualisation d'information selon Plaisant (2004).

L'évaluation effectuée avec l'aide d'utilisateurs permet de vérifier et de valider les hypothèses de conception et de faire des comparaisons entre les différents systèmes et techniques. Dans le domaine de l'interaction homme-machine, il existe de nombreuses méthodes et procédures qui peuvent nous aider à évaluer notre travail Andrews (2006), Carpendale (2008), Ellis et Dix (2006), Mazza et Berre (2007). La difficulté de l'évaluation des visualisations est liée à la nature de la connaissance du domaine. Dans notre cas, on n'a pas ce problème puisque notre visualisation s'adresse au grand public.

Hoeber (2009) propose un modèle d'évaluation et de raffinement pour l'étude systématique et l'amélioration des interfaces de recherche web (figure 5.2) et qui peut s'appliquer à d'autres systèmes de visualisation.

Ce modèle commence avec la conception et le développement d'un prototype préliminaire du système. Le deuxième niveau d'évaluation est l'étude en laboratoire. Ceux-ci sont effectués sur le prototype raffiné.

Bien qu'il puisse être nécessaire dans certains cas de procéder à des expériences de laboratoire au niveau des composants, la preuve de l'utilité des techniques employées viendra de l'évaluation de l'ensemble du système. Les opinions subjectives des participants sont également une mesure importante. Le principal résultat est la preuve subjective quant à l'efficacité du système et la satisfaction des participants. Suite à ce modèle, les chercheurs peuvent évaluer et affiner les aspects de l'interface d'une manière structurée, conduisant à une version bêta stable et entièrement opérationnelle du système appropriée pour les évaluations du monde réel.

Le principal inconvénient de ce modèle est le temps et l'effort qu'il faudrait pour l'appliquer à un projet de recherche ou de l'industrie. Cependant, l'avantage de ce modèle est qu'il permet aux chercheurs de déterminer la valeur de la technique proposée d'une manière progressive, en augmentant la qualité et l'exhaustivité de la mise en œuvre du prototype en cours de route.

Dans cette section, nous nous plaçons au niveau de l'évaluation restreinte du modèle proposé dans Hoeber (2009) (figure 5.2). Néanmoins, nous allons rappeler brièvement les étapes déjà réalisées et qui concordent avec les étapes proposées dans ce modèle.

Nous avons analysé les visualisations existantes dans le domaine de la météo (section 2.8). Nous avons conçu un prototype préliminaire. En nous basant sur les techniques de visualisation, de perceptions et d'interactivité, nous avons raffiné ce prototype (chapitre 4). Le prototype reflète les approches de personnalisation que nous proposons dans cette thèse. La plus grande partie du travail consistait à expérimenter, analyser et améliorer les méthodes proposées. En juin 2013, nous avons fait une démonstration du prototype au centre météorologique canadien pour les experts d'EC. Suite à cette démonstration, nous avons pris en considération les remarques des experts et développé une version bêta.

5.3.1 Scénarios

Revenons à l'étape de l'évaluation restreinte. Pour réaliser cette évaluation, nous devons préparer un scénario d'évaluation. Lam et al. (2012) présentent sept scénarios d'évaluation pour différents « types » ou « catégories » d'évaluation, nous en allons suivre deux :

l'évaluation du rendement de l'utilisateur : mesurer scientifiquement à l'aide des valeurs le rendement des utilisateurs dans l'accomplissement des tâches de l'évaluation.

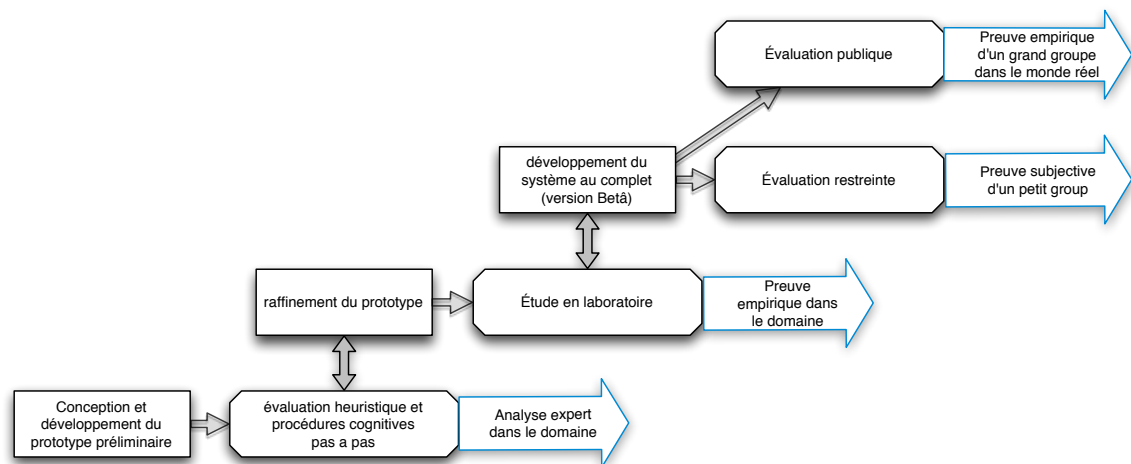


Figure 5.2 – Modèle d'évaluation et de raffinement, dans lequel les méthodes d'évaluation conduisent à poursuivre le raffinement et le développement du système de visualisation de l'information. Le résultat final est la preuve empirique d'un grand groupe de véritables utilisateurs. D'après (Hoeber (2009))

l'évaluation de l'expérience utilisateur : tenir compte de l'avis des utilisateurs sur la visualisation.

Même si les évaluateurs peuvent enregistrer des performances et des rétroactions d'un utilisateur, le but est de capturer une vue plus globale de l'expérience utilisateur. En revanche, l'évaluation doit se concentrer sur la visualisation elle-même, dans le but de tester les décisions de conception, d'explorer un espace de conception repère contre les systèmes existants, ou de découvrir les problèmes d'utilisation.

5.3.2 Évaluation du rendement de l'utilisateur

Objectifs et résultats : la performance de l'utilisateur est principalement mesurée en termes de paramètres objectivement mesurables tels que le temps et le taux d'erreur, mais il est également possible de mesurer la performance subjective telle que la qualité du travail aussi longtemps que les mesures peuvent être évaluées objectivement. Les indicateurs les plus couramment utilisés sont le délai d'exécution

et la précision de la tâche. Les sorties sont généralement des valeurs numériques analysées à l'aide des statistiques descriptives (comme la moyenne, la médiane, les écarts-types et intervalles de confiance) et modélisées par des méthodes telles que l'analyse de variance (Anova) pour partitionner la variance observée dans les composants.

Questions d'évaluation : les questions posées à l'aide des méthodes d'évaluation du rendement de l'utilisateur sont généralement précises et déterminées avant le début de l'évaluation. Il existe essentiellement deux types de questions :

Quelles sont les limites de la perception visuelle humaine et cognitive pour des types spécifiques de codage visuel ou techniques d'interaction ?

Comment peut-on comparer une visualisation ou technique d'interaction à une autre telle que mesurée par la performance humaine ?

Nous avons préparé quinze questions que les sujets auront à répondre à partir de la visualisation pour cette expérience (voir figure 5.3). Le choix des questions cherchait à évaluer toutes les parties et les paramètres de la visualisation. Nous avons ordonné les questions pour que les utilisateurs soient plus à l'aise. Deux choix se présentaient pour regrouper les questions.

Regroupement selon l'endroit où se trouvent les réponses attendues (condition courante, graphique de prévision ou description textuelle). L'inconvénient de ce type d'ordonnancement est que l'information peut être présentée dans plus d'une partie de la visualisation, mais l'utilisateur privilégiera celle dans la partie où il a trouvé la réponse aux questions précédentes même si ce n'est pas la plus appropriée pour répondre à la question.

Sachant que les sujets répondront aux mêmes questions pour MeteoVis et pour le bulletin d'EC, choisir le regroupement sur la base de l'endroit où se trouve la réponse pourrait être bénéfique pour une visualisation par rapport à l'autre. Nous avons plutôt

choisi de regrouper les questions autour de paramètres. Les questions auxquelles les utilisateurs ont répondu pour cette expérience sont ceux dans la figure 5.3.

1. Quelle est la température actuelle ?
2. Quelle est la température maximum prévue pour les quatre prochains jours ?
3. La température sera-t-elle à la hausse ou à la baisse ?
4. Donner la température (en degrés) pour trois prochaines heures ?
5. Y a-t-il précipitation actuellement ? Si oui, de quel type ?
6. Y a-t-il précipitation prévue pour les quatre prochains jours ?
7. Quelle précipitation est prévue pour demain après midi ?
8. Pour quel jour et heure la prochaine précipitation est elle prévue ?
9. Quelle est l'accumulation totale prévue pour la prochaine précipitation ?
10. Est ce qu'il y aura une tempête cette semaine ?
11. Quelles sont la direction et la vitesse du vent actuellement ?
12. Quelle vitesse du vent est prévue pour après demain matin ?
13. Quelle sera la journée la moins venteuse ?
14. Quelle sera la journée la plus venteuse ?
15. Lister les journées ensoleillées pour les quatre prochains jours ?

Figure 5.3 – questionnaire pour l'évaluation de la visualisation. Les questions sont regroupées selon les paramètres en question. Les quatre premières au sujet de la température. Les questions de 5 à 10 traitent de la précipitation et l'accumulation. Les questions de 11 à 14 sont à propos du vent. La dernière question pour la couverture nuageuse.

Le deuxième objectif de notre évaluation est de comparer notre système avec d'autres avec lesquels les participants effectuent les mêmes tâches. L'efficacité de l'interface est généralement définie par des mesures objectives telles que le temps d'obtention des réponses et leur exactitude.

Nous avons réalisé cette expérience avec sept utilisateurs (voir tableau 5.III). Chaque utilisateur devant de répondre aux mêmes questions pour la visualisation générée par MeteoVis et pour le bulletin météorologique sur le site web d'EC² (section 2.8.1). Pour

2. http://meteo.gc.ca/canada_f.html

éviter que l'utilisateur mémorise les réponses et par conséquent le temps de réponse pour la deuxième expérience soit plus rapide, l'utilisateur réalise l'expérience pour deux villes canadiennes différentes. En plus, pour plus de transparence et de précision la moitié des utilisateurs ont réalisé l'expérience d'abord avec le site EC puis pour notre visualisation et l'inverse pour l'autre moitié.

5.4 Évaluation de l'expérience utilisateur

Objectifs et résultats : L'évaluation de l'expérience utilisateur cherche à comprendre comment les gens réagissent à une visualisation. L'objectif est de comprendre dans quelle mesure la visualisation prend en charge les tâches prévues comme on le voit dans les yeux des participants et à la sonde pour les besoins et les préférences. Ce type d'évaluation produit des résultats subjectifs de ce qui est observé, recueillis ou mesurés. C'est le résultat des réponses subjectives des utilisateurs.

Les mesures comprennent la satisfaction, la confiance et les fonctionnalités aimées ou détestées. Les données recueillies dans cette étude peuvent nous aider à découvrir les lacunes en termes de fonctionnalités et les limitations dans la façon dont l'interface ou la visualisation est conçue et découvrir des directions prometteuses pour améliorer MeteoVis .

Questions d'évaluation : Qu'est-ce que les utilisateurs pensent de la visualisation ?

Plus spécifiquement, quelles sont les caractéristiques considérées utiles ?

Quelles sont les fonctionnalités manquantes ?

Comment les fonctionnalités peuvent-elles être retravaillées pour améliorer le processus de travail pris en charge ?

Y a-t-il des limites du système actuel qui entraverait son adoption ?

Est-ce que l'outil est compréhensible et peut-il être appris ?

5.5 Interprétation et amélioration

Le moyen de temps de réponse est 25% plus rapide pour la visualisation générée par MeteoVis comparé au bulletin météorologique préparé par EC. Une moyenne de 10,05s par question pour MeteoVis pour une moyenne de 13,41s pour le site d'EC.

Le contenu de notre visualisation est beaucoup plus riche que celui d'EC. Pour les mêmes questions, les utilisateurs n'ont pas trouvé de réponse pour 31% des questions en consultant le site d'EC. Toutefois, ils ont pu avoir la bonne réponse pour toutes les questions en consultant MeteoVis .

Les utilisateurs étaient très rapides (<5s) pour répondre aux questions 1,3,5 et 6. Pour la première question, c'est tout à fait normal puisque c'est l'information la plus recherchée pour chaque personne qui consulte la météo. Ajoutant à cela l'emplacement habituel où se trouve cette information dans la majorité des sites de météo. Ici l'utilisateur se base sur ses connaissances préalables que nous avons vues dans la section (2.8). Pour les questions 3, 5 et 6, l'information est présentée de manière à être détectée très rapidement et avec précision par le système visuel de bas niveau. On peut attribuer leur traitement cognitif à la première étape de l'attention visuelle, le traitement préattentif.

La température est présentée sous forme de petits cercles formant une courbe donnant directement l'allure (la tendance) de la température pour la prochaine période. Ce type de présentation ne demande presque pas de temps de traitement pour interpréter la tendance. Cette forme de présentation est détectée au moment du traitement préattentif.

La précipitation est représentée par des gouttes d'eau ou des flocons de neige en mouvement dans la visualisation. Cette présentation a facilité les réponses aux questions 5 et 6. La réponse était sous forme booléenne. La présentation l'était aussi. Il suffisait de jeter un coup d'oeil à la visualisation pour vérifier s'il existe ou non des gouttes d'eau (pluie) et/ou des flocons de neige. La réponse pour le type de la précipitation était sous forme booléenne et encore une fois la présentation l'était aussi.

Les utilisateurs ont eu plus de difficulté avec les questions 4,12 et 13 (plus de 15s). Même si la partie de la condition courante contient une section appelée « condition des prochaines heures », tous les utilisateurs ont eu tendance à pointer le curseur sur les trois premiers points de la visualisation graphique représentant la température pour connaître la température des trois prochaines heures. Ce qui explique le temps de réponse.

Nous avons délibérément posé des questions sur tous les paramètres. Nous savions que certains paramètres ne seraient pas affichés (visualisation personnalisée). Nous voulions évaluer le niveau d'interactivité des utilisateurs pour explorer d'autres paramètres de la visualisation. Le vent n'était présenté dans la visualisation. Pour trouver l'information concernant la prévision pour le vent, il fallait cocher ce paramètre dans le menu. Ce qui explique le temps de réponse pour la douzième question.

Tableau 5.III – Résultat de l'évaluation. Ui représente les utilisateurs. Chaque utilisateur a répondu aux 15 questions pour MeteoVis et pour le site web d'EC et a été chronométré pour chaque question. La dernière colonne représente la moyenne de temps de réponse pour chaque question. Les quatre premières questions concernent la température. Les questions de 5 à 10 concernent la précipitation et l'accumulation. Les questions de 11 à 14 concernent le vent. La quinzième question est à propos de la couverture nuageuse

	U1		U2		U3		U4		U5		U6		U7		Moyenne	
	VIS	EC	VIS	EC	EC	VIS	EC	VIS	EC	VIS	EC	VIS	EC	VIS	EC	VIS
1	1,0	0,4	4,8	1,5	12,1	6,8	1,7	1,5	0,4	0,8	1,4	1,4	2,7	0,3	2,9	2,4
2	14,8	3,4	25,3	14,5	50,3	12,6	10,6	11,3	10,7	6,4	5,7	15,8	7,6	15,3	14,7	14,5
3	2,8	0,4	3,2	4,8	22,7	9,8	10,3	7,0	3,6	0,4	9,5	2,8	16,3	9,0	9,7	5,0
4	18,0	30,2	56,1	21,0	60,7	15,9	34,3	19,5	24,9	9,9	25,2	28,9	54,8	5,6	35,9	22,0
5	1,3	3,1	0,4	27,0	29,5	10,1	6,1	3,2	20,6	9,6	10,6	6,1	15,1	3,6	16,0	4,9
6	0,5	0,6	0,3	3,1	33,3	4,2	1,5	4,6	12,8	7,8	4,0	1,9	5,2	14,0	8,6	4,8
7	4,2	5,6	7,3	9,4	35,3	15,0	3,9	4,5	6,2	0,7	15,4	6,7	18,7	14,3	13,5	7,5
8	8,8	15,6	6,4	13,5	13,4	3,2	36,1	28,8	11,4	12,4	3,1	9,5	3,3	7,5	13,8	10,9
9	13,5	18,0	1,1	23,3	18,1	19,4	68,4	2,6	23,6	1,7	16,3	17,1	36,4	3,4	29,1	8,4
10	0,5	13,9	14,0	8,4	10,2	10,4	14,9	22,0	8,6	6,4	8,2	0,4	0,4	3,4	9,2	8,2
11	7,5	5,3	20,7	18,9	15,2	4,8	8,7	1,0	14,8	13,3	1,9	23,0	9,3	3,4	10,6	10,5
12	13,4	6,8	4,7	15,1	19,4	25,1	9,4	26,1	19,4	7,2	10,4	25,0	18,1	6,8	14,1	15,5
13	5,3	12,8	10,3	1,9	30,2	6,4	9,8	20,5	6,5	10,2	10,0	53,4	23,4	0,9	13,5	15,3
14	24,0	1,7	10,7	1,7	7,3	5,1	4,7	8,6	3,4	1,0	11,1	21,8	1,3	1,2	4,5	10,3
15	0,3	1,5	3,5	9,8	7,0	27,1	10,0	17,8	2,3	2,3	2,6	18,4	3,5	5,6	5,2	10,7
	116,0	119,1	168,8	173,9	364,7	175,9	230,3	179,0	169,2	90,1	135,4	232,2	216,1	94,3	201,2	150,9

5.5.1 Seuil de décision

La perception d'une visualisation dépend de son contenu. Plus le contenu se rapproche des besoins de l'utilisateur plus la perception est facilitée. À la section 3.3 nous avons utilisé un seuil pour décider d'afficher ou non le paramètre étudié dans la visualisation. Ce seuil a été fixé arbitrairement au début des expérimentations. Pour décider de la robustesse de ce seuil, nous analysons sa sensibilité par rapport à l'interaction des utilisateurs avec la visualisation. Suite à cette interaction MeteoVis apprend au fur et à mesure les valeurs de ce seuil pour chaque paramètre selon le profil de l'utilisateur. Nous enregistrons les choix des utilisateurs pour les paramètres de la visualisation avec les informations de chaque utilisateur. Sachant que MeteoVis permet aux utilisateurs de modifier le contenu de la visualisation, nous considérons que si l'utilisateur modifie le contenu de la visualisation c'est que nous devons prendre en considération ce changement pour les prochaines visualisations. Si un paramètre n'est pas affiché, c'est parce que la valeur de ce paramètre dans le calcul des besoins et des préférences est inférieure au seuil de décision pour ce paramètre. Nous varions la valeur des seuils pour mieux répondre aux besoins des utilisateurs.

Pour évaluer le contenu de la visualisation, on se base sur le taux de satisfaction des utilisateurs. Les paramètres liés à la localisation influent beaucoup sur le résultat du clustering et par conséquent influent sur le contenu de la visualisation. Cela signifie que les participants à l'évaluation du contenu devraient être en grand nombre et dans des villes différentes un peu partout au Canada. Cette description de participants correspond à la dernière étape (évaluation publique) du modèle d'évaluation de Hoeber (2009). Sachant que MeteoVis améliore le contenu affiché sans intervention humaine en variant les valeurs des seuils de décision, l'évaluation du contenu pour la version actuelle du prototype n'a pas été réalisée dans cette thèse.

5.6 Conclusion

Nous générons une visualisation personnalisée selon le profil de l'utilisateur. Nous avons suivi les méthodes de perception et d'interactivité étudiées au chapitre 2. Nous avons mis au point des méthodes de personnalisation basées sur le profil de l'utilisateur. Le clustering a été utilisé pour dégager les besoins et les préférences selon les utilisateurs similaires qui ont utilisé MeteoVis dans le passé et sont sauvegardé dans une base de données contenant respectivement leurs profils et leurs préférences. Néanmoins, nous voulons améliorer cette visualisation. Pour évaluer la visualisation, nous avons analysé le taux de satisfaction des utilisateurs en nous basant sur leurs rétroactions. Dans ce chapitre, nous avons évalué le rendement des utilisateurs en étudiant leurs temps de réponse et l'exactitude des réponses. Nous avons aussi évalué l'expérience utilisateur pour comprendre dans quelle mesure la visualisation permet aux utilisateurs de réaliser les tâches prévues. MeteoVis apprend les habitudes des utilisateurs. Chaque nouvelle visualisation générée doit être meilleure que les précédentes au moins du point de vue des besoins et des préférences des utilisateurs. La variation du seuil de décision permet aux utilisateurs d'avoir un meilleur résultat. MeteoVis augmente ou diminue le seuil de décision automatiquement selon l'interactivité des utilisateurs.

CHAPITRE 6

CONCLUSION ET TRAVAUX FUTURS

Le travail présenté dans cette thèse touche à deux domaines : la construction d'un profil utilisateur, ses besoins et ses préférences et ce qu'on peut lui offrir ainsi que le domaine de la visualisation de l'information.

Nous avons proposé une approche pour modéliser une visualisation personnalisée en quatre étapes Mouine et Lapalme (2013a).

La première consiste en la compréhension, le nettoyage et l'extraction des données. Comme la visualisation peut être personnalisée en fonction de l'utilisateur. Nous avons débuté par une analyse des visualisations existantes en nous basant sur des principes et techniques cognitifs et perceptuels. Cette analyse nous a permis de déduire des règles de conception d'une visualisation dans le domaine de la météo.

Une grande partie de la problématique est la détection du profil utilisateur. Une des contributions de notre travail est l'approche que nous avons proposée pour détecter les besoins et les préférences d'un utilisateur presque inconnu en utilisant l'apprentissage non supervisé. L'innovation dans notre approche est que nous ne touchons pas à la vie privée des utilisateurs en fouillant dans leurs habitudes ou bien leurs comportements sur d'autres sites pour connaître leurs besoins. Nous nous basons sur le couple (profil, préférences) des utilisateurs antérieurs du système. Cette information est sauvegardée anonymement et sert au calcul des besoins et des préférences de l'utilisateur actuel.

La deuxième étape consiste à détecter automatiquement le profil de l'utilisateur et à prédire ses besoins et ses préférences. Nous nous sommes inspiré des méthodes d'aide à la décision multicritère pour fixer arbitrairement un seuil minimum des valeurs de préférences qui seront prises en compte dans la visualisation. Nous en avons analysé la sensibilité en le variant pour avoir des valeurs robustes vis-à-vis les préférences des

utilisateurs tel que rapporté dans Mouine et Lapalme (2013b). Nous avons fixé un seuil minimum pour décider des paramètres qui seront utilisés dans la visualisation. Nous utilisons le même seuil pour évaluer le niveau de satisfaction de l'utilisateur. En utilisant les données de la première étape et compte tenu des besoins et des préférences prédit.

La troisième étape crée une visualisation qui est évaluée et améliorée dans la dernière étape. Nous avons appliqué cette approche pour générer des bulletins météorologiques personnalisés pour les utilisateurs du site d'EC. La visualisation résultant prend en compte le profil de l'utilisateur actuel dont les besoins et les préférences sont basées sur les anciens utilisateurs ayant le même profil.

Nous avons préparé un prototype de MeteoVis que nous avons mis en ligne sur le serveur de l'Université de Montréal. Lorsque l'utilisateur se connecte à notre site web, MeteoVis détecte son emplacement grâce à son adresse IP. Le module de clustering est appelé pour calculer les paramètres qui seront affichés dans la visualisation. MeteoVis génère alors une visualisation personnalisée pour cet utilisateur en se basant sur les choix et les préférences des utilisateurs similaires.

Nous avons réalisé une évaluation restreinte basée sur des approches scientifiques pour avoir une rétroaction quantitative et qualitative de la part des utilisateurs. Nous avons préparé deux scénarios d'évaluation : dans le premier, nous avons comparé le rendement de l'utilisateur en utilisant notre système par rapport au système actuel d'EC. Une fois l'utilisateur avait répondu au questionnaire, nous lui demandions son avis sur la visualisation générée par MeteoVis. Les résultats de cette évaluation montrent que les visualisations générées par MeteoVis sont meilleures que les bulletins actuels préparés par EC. L'analyse des résultats de l'évaluation nous a permis d'améliorer plusieurs points et d'en prévoir d'autres.

6.1 Travaux futurs

Au début de cette thèse, nous avons plusieurs défis à atteindre et plusieurs problèmes à résoudre. Nous croyons que nous avons résolu le problème principal de notre travail en proposant une approche pour générer une visualisation personnalisée. Néanmoins, nous pensons que si nous avions eu plus de temps et de ressources, nous aurions pu traiter d'autres angles. Nous proposons dans cette section quelques perspectives et travaux futurs.

6.1.1 Évaluation publique

Selon la méthode proposée par Hoeber (2009) utilisée dans la section 5.3, la dernière étape est l'évaluation publique aussi appelé « études longitudinales » pour permettre aux participants d'utiliser le système à l'étude. Dans le cas de notre outil (MeteoVis) les participants sont la population canadienne en générale. Comme, il n'est pas nécessaire d'observer activement ou surveiller les participants (à la différence de l'évaluation restreinte que nous avons réalisée dans le chapitre précédent), on peut soutenir un plus grand nombre de participants dans une seule étude.

Les données de l'étude longitudinale pourraient être collectées en analysant l'interactivité réalisée par les utilisateurs sur la visualisation et l'évolution du trafic par rapport à l'avancement. Dans la section 3.5.1, nous avons montré comment MeteoVis ajuste le seuil de décision pour améliorer le contenu de la visualisation. L'analyse des résultats de ce type d'évaluation confirmerait ces hypothèses.

Les études longitudinales sont particulièrement bien adaptées à l'évaluation de l'interface de MeteoVis, car la consultation de la météo est une activité sur le lieu de travail, à l'école, à la maison, et à différents moments de la journée. Ces études longitudinales fournissent des informations précieuses sur l'utilité de l'interaction. Les études longitudinales capturent les activités et les impressions du système lors de l'utilisation d'un

utilisateur, avec un niveau modéré de contrôle, en termes de mesure des activités et des réactions subjectives, et le choix des participants. Les études ne sont pas contraintes par un lieu ou un moment précis, ce qui les rend plus souples que d'autres méthodes d'évaluation. La preuve empirique est capturée.

Dans notre cas, pour réaliser cette évaluation il nous faut l'accord d'EC pour héberger l'application sur ses serveurs ou bien sur un serveur puissant pouvant supporter le trafic résultant. Si nous avons cette opportunité, nous pourrions déduire un taux de satisfaction des utilisateurs par l'analyse de l'évolution du trafic. L'accroissement du nombre d'utilisateurs reflèterait une opinion générale positive. Nous pouvons aussi utiliser des méthodes statistiques (e.g sondage) pour une évaluation plus approfondie. La possibilité de laisser des commentaires à propos de la visualisation et aussi une méthode que nous pourrions utiliser pour améliorer MeteoVis dépendamment des avis des utilisateurs.

6.1.2 Génération de texte

Nous générons le texte des bulletins situés au bas de la page web en anglais et en français, les deux langues officielles du Canada. Pour l'instant nous avons utilisé les gabarits de phrases pour générer ce texte. Une autre approche pourrait être utilisée pour améliorer cette partie. Nous proposons l'utilisation de la bibliothèque Java SimpleNLG-ENFR. C'est une version adaptée par Pierre-Luc Vaudry à partir de SimpleNLG v4.2 (Gatt et Reiter, 2009) qui permet de réaliser du texte en français et en anglais. La figure 6.1 est un exemple de la génération de la phrase : « *On prévoit des vents forts d'ouest à 90 km/h ou plus accompagnés de pluie.* » qui suit la méthode décrite dans Reiter et Dale (2000) et qui est générée à l'aide de SimpleNLG-ENFR.

Nous créerions un module, semblable au module d'extraction de données (figure 4.1), pour générer la description textuelle de la météo. La génération actuelle est réalisée par des gabarits de phrases que MeteoVis choisit selon les valeurs des données mé-

Figure 6.1 – Exemple de code Java de SimpleNLG-ENFR permettant de générer la phrase « On prévoit des vents forts d’ouest à 90 km/h ou plus accompagnés de pluie. » Nous donnons en paramètres les valeurs de la vitesse et la direction du vent et le type de précipitation. On prépare la description du vent auquel on ajoute la précipitation pour réaliser la phrase. Nous remercions Pierre Luc Vaudry pour son aide dans la réalisation de cet exemple.

```
public static String ventPluie(NLGFactory nlgF,
    int vitesseMin,int vitesseMax,String direction,
    String qualite, String typePrecipitation){

    // description du vent
    NPPhraseSpec vent = nlgF.createNounPhrase("un", "vent");
    if (qualite!=null){
        vent.setFeature(Feature.NUMBER, NumberAgreement.PLURAL);
        vent.addModifier(qualite);
    }
    vent.addModifier(nlgF.createPrepositionPhrase("de",direction));
    String vitesse=vitesseMin+" km/heure";
    if(vitesseMax>vitesseMin)vitesse+=" ou plus";
    vent.addModifier(nlgF.createPrepositionPhrase("à",vitesse));

    // ajout des précipitations
    if(typePrecipitation!=null){
        NLGElement accompagner=
            nlgF.createWord("accompagner",LexicalCategory.VERB);
        accompagner.setFeature(Feature.FORM,Form.PAST_PARTICIPLE);
        AdjPhraseSpec ap=nlgF.createAdjectivePhrase(accompagner);
        ap.addComplement(nlgF.createPrepositionPhrase("de",
            typePrecipitation));
        vent.addModifier(ap);
    }

    // réalisation de la phrase
    return new Realiser()
        .realiseSentence(nlgF.createClause("on","prévoir",vent));
}
```

téorologiques. MeteoVis ferait appel à ce module (SimpleNLG-ENFR) en lui passant ces valeurs pour générer des phrases entières. La communication entre MeteoVis et SimpleNLG-ENFR se ferait à l'aide d'un code JavaScript/AJAX.

6.1.3 Application à d'autres domaines

Notre approche personnalise selon le profil de l'utilisateur l'affichage d'une grande masse d'information évoluant dans le temps. Le but principal de notre travail est de permettre à l'utilisateur d'extraire facilement de la visualisation les informations dont il a besoin. Nous proposons deux autres domaines d'application de notre approche.

6.1.4 Médecine

Les membres du personnel clinique (médecins, techniciens, infirmières, étudiants, gestionnaires) risquent d'être submergés par la grande quantité de données des patients (analyses, diagnostics, imagerie médicales...) si les applications informatiques ne prennent pas suffisamment en compte la présentation et l'interaction efficace avec ces données Chittaro (2001).

Dans le domaine de la médecine, le médecin analyse les diagnostics et les symptômes du patient pour lui prescrire un traitement. Le jugement du médecin est basé sur ses compétences scientifiques et aussi sur des cas similaires traités dans le passé. Nous pensons que notre approche peut être d'une grande utilité aux médecins. On générerait une visualisation personnalisée au cas étudié pour chaque patient. Les paramètres de la visualisation selon notre approche seraient par exemple : les symptômes des patients, le traitement suivi et le résultat du traitement pour chacun.

Dans le cas de MeteoVis, c'est l'état et la prévision de la météo qui sont affichés. Les paramètres affichés dans la visualisation dépendent du profil de l'utilisateur. Dans le domaine de la médecine ce seraient les résultats des traitements suivis par les cas

(patients) similaires en fonction des symptômes et des diagnostics du patient.

Nous commencerons par l'application de l'algorithme de clustering pour regrouper les patients similaires. Le calcul du degré de similarité dans ce domaine peut s'avérer profitable. Nous rajouterons une variable ici permettant au médecin de décider d'augmenter ou de diminuer le seuil de degré de similarité pris en compte. Nous avons pondéré les critères du profil utilisateur pour MeteoVis dans le but de donner plus de poids pour les critères les plus importants, nous ferions de même pour les paramètres des diagnostics et des symptômes. La visualisation générée serait personnalisée au cas étudié. Le médecin (utilisateur) aurait la possibilité d'interagir avec la visualisation pour pouvoir agir sur tous les paramètres de la visualisation. Enfin, et comme dans le cas de MeteoVis le système apprendrait à modifier les valeurs des paramètres selon l'interactivité de l'utilisateur pour générer de meilleurs résultats dans le futur.

6.1.5 Finance

Nous présentons dans cette section comment appliquer notre approche au marché des changes ou aussi appelé « Forex ». Dans ce type de marché, les devises sont échangées. Les profits et les pertes sont dus aux changements de taux de change sensibles aux nouvelles économiques, financières et politiques.

Les négociants, « trader », doivent prédire les fluctuations du marché pour réussir des transactions profitables. Les informations pouvant les aider à prédire les fluctuations du marché leurs permettent de bâtir une opinion sur l'état futur des taux de devises.

La plupart des courtiers mettent à la disposition de leurs clients des calendriers économiques présentant les nouvelles en rapport avec le marché des devises. La figure 6.2 représente un exemple de calendrier économique, l'un des moyens le plus utilisé pour prédire les variations des taux.

Si on considère les habitudes de spéculation des « traders » comme étant leur profil,

on peut remarquer que notre approche s'applique facilement à ce type de domaine.

Comme pour MeteoVis la génération de la visualisation dans ce cas se baserait sur le profil de l'utilisateur. Les données affichées seraient celles du calendrier économique et leur impact prévu sur le marché. Le choix des données affichées et les devises concernées dépendrait des habitudes de l'utilisateur et du choix fait par les utilisateurs similaires.

La projection de notre approche à ce domaine est directe. Nous regrouperions les utilisateurs avec des habitudes de « trading » similaires. Comme dans le cas de MeteoVis, selon la similarité des profils nous calculerions les préférences de notre utilisateur pour générer une visualisation résumant les données du calendrier économique en se limitant aux données jugées importantes. L'utilisateur interagirait avec la visualisation de telle sorte qu'il peut demander plus ou moins de détails, changer la configuration de la visualisation, ajouter et/ou enlever des paramètres...

Figure 6.2 – Calendrier économique

DAILY WEEKLY MONTHLY View: Calendar List Sort By: Date Impact Filter ▼							
Date	Time ET	Country	Event	Impact	Actual	Forecast	Previous
March 21	10:00AM		US: Atlanta Fed Business Inflation Expectations		%	%	2.0%
March 21	8:30AM		CA: CPI		%	0.6%	0.3%
March 21	8:30AM		CA: Retail Sales		%	0.7%	-1.8%
March 21	5:30AM		GB: Public Sector Finances		Stgbillion	Stg7.7billion	Stg-6.4billion
March 20	4:30PM		US: Money Supply		\$B	\$B	\$-13.9B
March 20	4:30PM		US: Fed Balance Sheet		\$B	\$B	\$9.6B
March 20	10:30AM		US: EIA Natural Gas Report		bcf	bcf	-195bcf
March 20	10:00AM		US: Leading Indicators		%	0.3%	0.3%
March 20	10:00AM		US: Existing Home Sales		M	4.600M	4.62M
March 20	10:00AM		US: Philadelphia Fed Survey			3.0	-6.3
March 20	9:45AM		US: Bloomberg Consumer Comfort Index				-27.6
March 20	8:30AM		US: Jobless Claims		K	325K	315K
March 20	7:00AM		GB: CBI Industrial Trends Survey			6.0	3
March 20	4:30AM		CH: SNB Monetary Policy Assessment		bp	0bp	0bp
March 20	3:00AM		CH: Merchandise Trade Balance				2.6
March 20	3:00AM		DE: PPI		%	%	-0.1%
March 19	5:45PM		NZ: GDP		0.9%	0.9%	1.4%
March 19	2:00PM		US: FOMC Meeting Announcement		0 to 0.25%	0 to 0.25%	0 to 0.25%
March 19	2:00PM		US: FOMC Forecasts				
March 19	10:30AM		US: EIA Petroleum Status Report		5.9M barrels	6.1M barrels	6.2M barrels

6.2 Conclusion

La visualisation de l'information permet d'analyser les données pour comprendre son contenu. Dans cette thèse, nous avons proposé une approche complète pour générer une visualisation ayant un contenu personnalisé selon les préférences d'un utilisateur quasi inconnu. Le profilage de l'utilisateur selon notre approche se concentre sur la prédiction de ses besoins et préférences sans pour autant toucher à sa vie personnelle en se basant sur le comportement de ses semblables. Dans cette thèse, nous avons utilisé des prédictions de très haute qualité corrigée à la main par des météorologues. Néanmoins, ces prédictions sont de faible résolution (la géographique concernée). EC produit aussi des fichiers appelés GRIB contenant des prédictions à très haute résolution. Il serait intéressant dans le futur de combiner les deux pour interpoler les observations sur une grille de haute résolution pour fournir aux l'utilisateurs une prévision plus personnalisée.

BIBLIOGRAPHIE

- M. Agrawala, W. Li et F. Berthouzoz. Design principles for visual communication. *Communications of the ACM*, 54(4):60–69, 2011.
- Keith Andrews. Evaluating information visualisations. Dans *Proceedings of the 2006 AVI workshop on beyond time and errors : novel evaluation methods for information visualization*, pages 1–5. ACM, 2006.
- Hal Berghel. Hijacking the web. *Communications of the ACM*, 45(4):23–27, 2002.
- J.P. Brans, P. Vincke et B. Mareschal. How to select and how to rank projects : The promethee method. *european journal of operational research*, 24(2):228–238, 1986.
- I. Cadez, D. Heckerman, C. Meek, P. Smyth et S. White. Visualization of navigation patterns on a web site using model-based clustering. Dans *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 280–284. ACM, 2000.
- Sheelagh Carpendale. Evaluating information visualizations. Dans *Information Visualization*, pages 19–45. Springer, 2008.
- Chaomei Chen. Top 10 unsolved information visualization problems. *Computer Graphics and Applications, IEEE*, 25(4):12–16, 2005.
- Luca Chittaro. Information visualization and its application to medicine. *Artificial Intelligence in Medicine*, 22(2):81–88, 2001.
- W.S. Cleveland. *The elements of graphing data*. AT&T Bell Laboratories, 1994. ISBN 9780963488411.

- M. Corio et G. Lapalme. Generation of texts for information graphics. Dans *Proceedings of the 7th European Workshop on Natural Language Generation EWNLG'99*, pages 49–58. Citeseer, 1999.
- S. Dasgupta. The hardness of k-means clustering. *Techn. Rep. no. CS-2007-0890 (Univ. California, 2007)*, 2008.
- Geoffrey Ellis et Alan Dix. An explorative analysis of user evaluation studies in information visualisation. Dans *Proceedings of the 2006 AVI workshop on BEyond time and errors : novel evaluation methods for information visualization*, pages 1–7. ACM, 2006.
- Niklas Elmqvist et Ji Soo Yi. Patterns for visualization evaluation. Dans *Proceedings of the 2012 BELIV Workshop : Beyond Time and Errors-Novel Evaluation Methods for Visualization*, page 12. ACM, 2012.
- Massimo Fasciano. *Génération intégrée de textes et des graphiques statistiques*. Thèse de doctorat, Université de Montréal, 1996.
- A. Gatt et E. Reiter. Simplenlg : A realisation engine for practical applications. Dans *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93. Association for Computational Linguistics, 2009.
- Joseph Goldberg et Jonathan Helfman. Eye tracking for visualization evaluation : Reading values on linear versus radial graphs. *Information Visualization*, 10(3):182–195, 2011.
- Georges Grinstein, Alfred Kobsa, Catherine Plaisant, Ben Shneiderman et John T. Stasko. Which comes first, usability or utility ? Dans *Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, VIS '03, pages 112–, Washington, DC, USA, 2003.

- IEEE Computer Society. ISBN 0-7695-2030-8. URL <http://dx.doi.org/10.1109/VISUAL.2003.1250426>.
- R.L. Harris. *Information graphics : A comprehensive illustrated reference*. 978-0-19-513532-9. Oxford University Press, USA, 1999.
- J.A. Hartigan et M.A. Wong. Algorithm as 136 : A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.
- Orland Hoeber. User evaluation methods for visual web search interfaces. Dans *Information Visualisation, 2009 13th International Conference*, pages 139–145. IEEE, 2009.
- Amir M Hormozi. Cookies and privacy. *EDPACS*, 32(9):1–13, 2005.
- D A Keim. Information visualization and visual data mining. *IEEE transactions on visualization and computer graphics*, 8(1):1–8, 2002. ISSN 1077-2626.
- Robert Kosara, David H Laidlaw, Colin Ware, Christopher G Healey et Victoria Interrante. User studies : why, how, and when ? *IEEE Computer Graphics and Applications*, 23(4):20–25, 2003.
- D. Kostadinov. Personnalisation de l’information et gestion des profils utilisateurs. *Mémoire de DEA PRiSM, Versailles*, 2003.
- David M Kristol. Http cookies : Standards, privacy, and politics. *ACM Transactions on Internet Technology (TOIT)*, 1(2):151–198, 2001.
- Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant et Sheelagh Carpendale. Empirical studies in information visualization : Seven scenarios. *Visualization and Computer Graphics, IEEE Transactions on*, 18(9):1520–1536, 2012.

- L.Y. Maystre, J. Pictet et J. Simos. *Méthodes multicritères ELECTRE : description, conseils pratiques et cas d'application à la gestion environnementale*, volume 8. PPUR, 1994.
- Riccardo Mazza et Alessandra Berre. Focus group methodology for evaluating information visualization techniques and tools. Dans *Information Visualization, 2007. IV'07. 11th International Conference*, pages 74–80. IEEE, 2007.
- Laura McCarthy et Dave Yates. The use of cookies in federal agency web sites : Privacy and recordkeeping issues. *Government Information Quarterly*, 27(3):231–237, 2010.
- Ann McNamara, Katerina Mania et Diego Gutierrez. Perception in graphics, visualization, virtual environments and animation. Dans *SIGGRAPH Asia 2011 Courses*, page 17. ACM, 2011.
- S.B. Mena. Une solution informatisée à l'analyse de sensibilité d'electre iii. *Biotechnol. Agron. Soc. Environ*, 5(1):31–35, 2001.
- V.O. Mittal, G. Carenini, J.D. Moore et S. Roth. Describing complex charts in natural language : A caption generation system. *Computational Linguistics*, 24(3):431–467, 1998.
- B. Mobasher, H. Dai, T. Luo et M. Nakagawa. Discovery and evaluation of aggregate usage profiles for web personalization. *Data mining and knowledge discovery*, 6(1): 61–82, 2002.
- M. Mouine et G. Lapalme. Using clustering to personalize visualization. Dans *Information Visualisation (IV), 2012 16th International Conference on*, pages 258–263. IEEE, 2012.

- M. Mouine et G. Lapalme. How to model a customized visualization. Dans *Information Visualisation (IV), 2013 17th International Conference*, pages 118–122. IEEE, 2013a.
- M. Mouine et G. Lapalme. Preference thresholds optimization by interactive variation. Dans *Advances in Artificial Intelligence*, pages 286–292. Springer, 2013b.
- Tien N Nguyen et Jun Zhang. A novel visualization model for web search results. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):981–988, 2006.
- D. Pineo et C. Ware. Data visualization optimization via. computational modeling of perception. *IEEE Transactions on Visualization and Computer Graphics*, 2011.
- Catherine Plaisant. The challenge of information visualization evaluation. Dans *Proceedings of the working conference on Advanced visual interfaces*, pages 109–116. ACM, 2004.
- E. Reiter et R. Dale. *Building natural language generation systems*. Cambridge Univ Pr, 2000.
- Naomi B Robbins. *Creating more effective graphs*. John Wiley & Sons, 2012.
- B. Roy. Classement et choix en présence de points de vue multiples (la méthode electre). *Riro*, 2(8):57–75, 1968.
- B. Roy. Electre iii : Un algorithme de classement fondé sur une représentation floue des préférences en présence de critères multiples. *Cahiers du CERO*, 20(1):3–24, 1978.
- Rajiv C Shah et Jay P Kesan. Recipes for cookies : how institutions shape communication technologies. *New Media and Society*, 11(3):315–336, 2009.
- S. Sripada, E. Reiter et I. Davy. Sumtime-mousam : Configurable marine weather forecast generator. *Expert Update*, 6(3):4–10, 2003.

- S.G. Sripada, E. Reiter, J. Hunter et J. Yu. Sumtime : Observations from ka for weather domain. Rapport technique, Citeseer, 2001.
- R Verret, D Vigneux, J Marcoux, R Parent, F Petrucci, C Landry, L Pelletier et G Hardy. Scribe 3.0 a product generator. Dans *Preprints 13th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, pages 2–7, 1997.
- C. Ware. *Information visualization : perception for design*. Morgan Kaufmann, 2012.
- Chris Weaver. Look before you link : Eye tracking in multiple coordinated view visualization. *BELIV 10 : BEyond time and errors : novel evaluation methods for Information Visualization*, page 2, 2010.
- Wojciech Wiza, Krzysztof Walczak et Wojciech Cellary. Periscope : a system for adaptive 3d visualization of search results. Dans *Proceedings of the ninth international conference on 3D Web technology*, pages 29–40. ACM, 2004.
- J.S. Yi, Y. ah Kang, J.T. Stasko et J.A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1224–1231, 2007.
- J. Yu. *SumTime-Turbine : a knowledge-based system to generate English textual summaries of gas turbine time series data*. Thèse de doctorat, University of Aberdeen., 2004.
- G. Zelazny. *Say it with charts : the executive's guide to visual communication*. 0-7863-0894-X. McGraw-Hill Companies, 2001.