

## **RELEVANSI HASIL PENCARIAN PADA MESIN PENCARI BERDASARKAN KEDEKATAN KATA MENGGUNAKAN ONTOLOGI**

### *Search Results Relevance of Search Engine Based On Using Word Proximity Ontology*

**Maskur**

Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang

Email : maskur.informatika@gmail.com

#### **ABSTRACT**

*Search on search engines now is search-based keyword matching which keywords matched with the words that exist in the content. It tends to produce content relevant to the context of the keywords entered. Users should make the selection of relevant search results page with search results submitted by keywords entered into search mesin. This research proposed a method to filter ontology, then the first step taken is to prepare ontology as a search reference material. The next step is to use keywords and search results obtained in the search engines look for proximity to the word ontology. In the process of looking for the proximity of words using the existing rules on ontology. If the process of seeking closeness is found it will be calculated how close the keywords with the search results. The system will display search results that have the proximity of keywords with the search results of search engines. Thus generated search results that are relevant to the keyword. Therefore, in this study a method is proposed for calculating the proximity of keywords with the search results on ontology. Performance measurement is done by comparing the results of the search engines with results filtering search results using the ontology. Results of the testing that has been conducted shows that the application of these methods produce results that are relevant to the keywords users. Relevance of search results with ontologies produce better search than the search performed by the search engine based on the proximity of keywords with the search results.*

*Keyword: relevance, keyword, important words, search engines, ontology.*

#### **ABSTRAK**

Pencarian pada mesin pencari yang ada sekarang adalah pencarian yang berbasis pencocokan kata kunci dimana kata kunci dicocokkan dengan kata-kata yang ada pada konten. Hal tersebut cenderung menghasilkan konten yang tidak relevan dengan konteks dari kata kunci yang dimasukkan. Pengguna harus melakukan pemilihan halaman hasil pencarian yang relevan dengan hasil pencarian yang diajukan berdasarkan kata kunci yang dimasukkan ke mesin pencari. Penelitian ini mengajukan sebuah metode untuk menyaring ontologi, maka langkah awal yang dilakukan adalah mempersiapkan ontologi sebagai bahan acuan pencarian. Langkah selanjutnya kata kunci yang digunakan dan hasil pencarian yang didapatkan pada mesin pencari dicari kedekatan kata pada ontologi. Didalam proses mencari kedekatan kata menggunakan aturan yang ada pada ontologi. Apabila proses mencari kedekatan sudah ditemukan maka akan dihitung seberapa dekat kata kunci dengan hasil pencarian. Sistem akan menampilkan hasil pencarian yang memiliki kedekatan kata kunci dengan hasil pencarian dari mesin pencari. Sehingga dihasilkan hasil pencarian yang relevan dengan kata kunci. Oleh karena itu, dalam penelitian ini suatu metode diusulkan untuk menghitung kedekatan kata kunci dengan hasil pencarian pada ontologi. Pengukuran kinerja dilakukan dengan cara membandingkan hasil pencarian dari mesin pencari dengan hasil penyaringan hasil pencarian yang menggunakan ontologi. Hasil pengujian yang telah dilakukan menunjukkan bahwa penerapan metode ini menghasilkan hasil pencarian yang relevan dengan kata kunci pengguna. Relevansi hasil pencarian dengan ontologi menghasilkan pencarian yang lebih baik dari pada pencarian yang dilakukan oleh mesin pencari berdasarkan kedekatan kata kunci dengan hasil pencarian.

Kata kunci : *Controller*, LED, Bertenaga Surya.

#### **PENDAHULUAN**

Perkembangan internet meningkat dengan sangat cepat. Bahkan internet sudah

menjadi bagian dari gaya hidup sehari-hari bagi beberapa kalangan. Hal ini karena penyebaran informasi dengan media internet sangat cepat tanpa ada batasan waktu dan tempat. Orang-orang di seluruh dunia dapat saling mempublikasikan sumber daya yang mereka miliki di internet sehingga informasi tersebar dimana-mana. Untuk mendapatkan informasi yang kita butuhkan di internet kita bisa menggunakan mesin pencari yang juga sudah banyak tersedia. Informasi yang dihasilkan oleh mesin pencari terkadang masih belum sesuai dengan yang diinginkan oleh pengguna. Mesin-mesin pencari tersebut melakukan pencarian berdasarkan kata kunci yang dimasukkan oleh pengguna selanjutnya mencocokkan kata kunci dengan konten-konten yang tersebar di internet. Hasil pencarian dengan menggunakan metode ini banyak memberikan hasil yang terkadang tidak sesuai dengan konteks informasi yang kita inginkan. Hal ini karena pada proses pencariannya tidak mempertimbangkan konteks dari informasi atau konten yang kita cari, hanya mencocokkan kata-kata yang ada pada suatu konten dengan kata kunci (Pramudiono, 2006).

Metode pencarian ini dengan cara melakukan pengindekan pada masing-masing id dari dokumen berdasarkan konteks dan *term*. Belum ada metadata yang standart yang dapat digunakan. Cara ini dianggap masih perlu penyempurnaan karena hasil pencarian dokumen yang memiliki hubungan dengan kata kunci masih belum relevan (Gupta & Sharma, 2010). Ontologi mampu menyelesaikan keterbatasan makna karena di dalam ontologi terdapat hubungan antar kata (konsep) yang dapat membedakan kata berdasarkan kegunaan dan makna. Dari penelitian sebelumnya diajukan metode pencarian untuk menyaring pencarian berbasis kata kunci pada mesin pencari menggunakan ontologi dengan tujuan untuk relevansi hasil pencarian.

## METODE PENELITIAN

### Mesin Pencari Berbasis Kata Kunci

Mesin pencari berbasis kata kunci adalah suatu metode pencarian berbasis kata kunci. Dimana ketika kita melakukan pencarian, kata kunci yang dimasukkan dan akan muncul hasil berdasarkan kata kunci tersebut. Hasil yang didapat seringkali tidak sesuai dengan yang diinginkan sehingga dapat menyebabkan kebanjiran informasi, waktu yang tidak efektif dan pengkonsumsian sumber daya yang berlebihan.

### Ontologi

Istilah ontologi sebenarnya berasal dari istilah filosofi "*ontology*" yang artinya sesuatu yang sesungguhnya ada dan bagaimana menggambarkannya. Dalam dunia komputer ontolog digunakan untuk menspesifikasikan suatu konseptualisasi. Dalam istilah lain ontologi dijelaskan sebagai suatu representasi dari domain pengetahuan tertentu yang berisi istilah-istilah dalam domain tersebut beserta hubungan antara istilah-istilah yang ada.

Ontologi saat ini banyak digunakan terutama untuk mendukung web semantik, yaitu teknologi web yang diarahkan dapat memahami makna suatu kata atau kalimat yang diberikan oleh pengguna. Membuat komputer mengerti seperti manusia adalah suatu hal yang sepertinya mustahil, namun visi ini teru diupayakan dengan menyediakan seperangkat alat sehingga membuat mesin atau komputer dengan mudah dapat memproses informasi dan mengerti informasi yang diinginkan oleh pengguna.

Tidak ada standar khusus untuk membangun suatu ontologi dan tidak ada justifikasi bahwa ontologi yang dikembangkan oleh seseorang adalah salah atau benar. Kualitas ontologi dapat dilihat dari aplikasi yang dibangun berdasarkan ontologi ini. Ketika aplikasi yang dibangun dapat memenuhikebutuhan pengguna dan

menjawab permasalahan yang ada maka ontologi yang digunakan termasuk ontologi yang berkualitas (Antoniou & Harmelen, 2004).

### Bahasa Ontologi

Ontologi sendiri mempunyai struktur bahasa yang formal (terdefinisi), agar dapat digunakan. Beberapa struktur bahasa yang menyusun ontology antara lain (Pramudiono, 2006):

- XML (*Extensible Markup Language*) : Struktur mirip HTML yang tag-nya dapat didefinisikan sendiri.
- XML *Schema* : Bahasa yang membatasi struktur yang didefinisikan pada dokumen XML.
- RDF (*Resource Description Framework*) : Model data untuk objek (“resources”) dan relasi diantaranya, menyediakan semantik yang sederhana untuk model data tersebut, dan data model ini dapat disajikan dalam sintaksis XML.
- RDF *Schema* : Adalah kosakata untuk menjelaskan *properties* dan *classes* dari sumber RDF, dengan sebuah *semantics* untuk hirarki penyamarataan dari *properties* dan *classes*.
- OWL (*Ontology Web Language*) : Menambahkan beberapa kosakata untuk menjelaskan *properties* dan *classes*, antara lain : relasi antara *classes* (misalkan *disjointness*), kardinalitas (misalkan ‘tepat satu’), *equality*, berbagai tipe dari *properties*, karakteristik dari *properties* (misalkan *symmetry*), menyebutkan satu persatu *classes*. Pembahasan mengenai beberapa struktur bahasa diatas terdapat pada sub bab tersendiri.

### OWL (*Ontology Web Language*)

*Web Ontology Language* (OWL) adalah suatu bahasa yang dapat digunakan

oleh aplikasi–aplikasi yang bukan sekedar menampilkan informasi tersebut pada manusia, melainkan juga yang perlu memproses isi informasi isi. Ontology sendiri dapat didefinisikan sebagai suatu cara untuk mendeskripsikan arti dan relasi dari istilah-istilah. Deskripsi tersebut berisi *classes*, *properties*, dan *instances*. Deskripsi ini dapat membantu sistem komputer dalam menggunakan istilah-istilah tersebut dengan cara yang lebih mudah.

Dengan menggunakan OWL, kita dapat menambah *vocabulary* tambahan disamping semantik formal yang telah dibuat sebelumnya menggunakan XML, RDF, dan RDF Schema. Hal ini sangat membantu penginterpretasian mesin yang lebih baik terhadap isi Web. Untuk mendeskripsikan *properties* dan *classes*, OWL menambahkan *vocabulary* seperti (Wicaksana, 2006) :

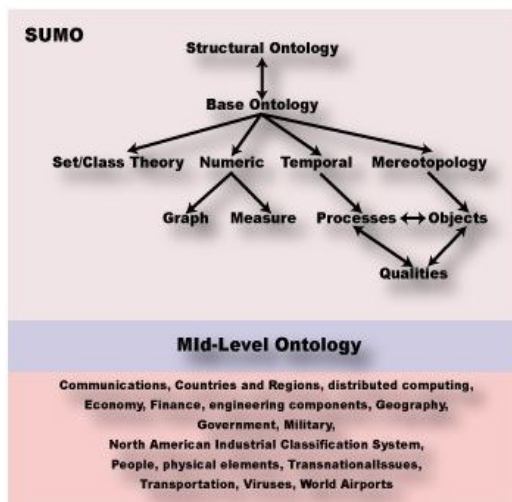
- “among others”.
- Relasi antar *classes* (misalnya: “*disjointness*”).
- Kardinalitas (misalnya: “*exactly one*”).
- Kesamaan (*equality*).
- Karakteristik *property* (misalnya: “*symmetry*”).
- *Enumerated classes*.

## HASIL DAN PEMBAHASAN

### Ontologi Dengan SUMO

SUMO adalah ontologi terdiri dari sekitar 1000 konsep dan 4000 definisi laporan. SUMO dikembangkan melalui berbagai tahap pengembangan dan eksperimentasi, yang membuatnya stabil dan matang sehingga untuk dianggap sebagai “standar” ontologi. Selain itu, SUMO telah dikembangkan dengan Ontologi *Mid-Level* (MILO), dan sejumlah domain ontologi yang memungkinkan cakupannya untuk berbagai aplikasi domain. Salah satu fitur menarik dari SUMO adalah bahwa berbagai fiturnya sub-ontologi adalah independen dan dapat digunakan sendiri atau dikombinasikan.

Ontologi *Mid-Level* (MILO) domainnya meliputi : komunikasi, negara dan wilayah, komputasi terdistribusi, ekonomi, keuangan, mobil dan komponen teknik, geografi, pemerintah, taksonomi, media, Militer (umum, perangkat, proses, orang), transportasi, virus, dunia bandara, dan senjata pemusnah massal (Zouaq, et al., 2009).



Gambar 1. Ontologi Dengan SUMO

## Pengumpulan dan Analisis Data

Dataset yang akan digunakan dalam penelitian ini adalah dokumen berbahasa Inggris. Proses pengumpulan data dilakukan dengan cara mengambil hasil pencarian dari mesin pencari menggunakan *google search engine Application Programming Interface (API)* dan disimpan dalam *database*. Data yang didapatkan meliputi *title, url* dan konten.

## Preprocessing

### Eliminasi Tag HTML

Eliminasi penghilangan Tag HTML yang dianggap tidak diperlukan dalam konten. Seluruh Tag HTML yang ada dalam konten akan dihilangkan. Eliminasi tag HTML berguna untuk kepentingan pada proses selanjutnya. Tag HTML tidak diperlukan sehingga harus dihilangkan dari dokumen.

## Eliminasi Tanda Baca

Eliminasi penghilangan tanda baca yang dianggap tidak diperlukan dalam konten. Seluruh tanda baca yang ada dalam konten akan dihilangkan. Eliminasi ini bertujuan agar proses selanjutnya. Semua tanda baca yang ada pada dokumen tidak berguna untuk kepentingan pada proses selanjutnya. Semua tanda baca yang ada pada dokumen tidak diperlukan.

## POS Tagging

Proses ini dilakukan untuk mendapatkan informasi *tag* dari tiap kata yang ada pada dataset. Sehingga dari dokumen yang diolah dapat diketahui mana kata kerja, kata benda dan kata sifat. Pada proses *POS Tagging* ini isi pada dokumen akan dicari kata-kata yang penting. Semua kata pada dokumen akan dikenali sebagai kata apa untuk memudahkan untuk proses selanjutnya.

## Stopword Removal

Eliminasi dari *stopword* yaitu penghilangan kata yang dianggap tidak diperlukan. Pada proses ini kata-kata yang dianggap tidak diperlukan akan dihilangkan. Semua kata yang dianggap tidak mempunyai kepentingan akan dihilangkan sehingga didapatkan kata-kata yang diperlukan.

## Pembobotan Term

Proses ini dilakukan untuk mendapatkan bobot *term* pada dokumen. Pembobotan *tf-idf (term frequency-invert document frequency)* adalah pembobotan yang sering digunakan dalam temu kembali informasi dan *text mining*. Pembobotan ini adalah pengukuran statistik yang digunakan untuk mengevaluasi pentingnya sebuah kata ke dokumen dalam suatu koleksi dokumen atau *corpus*. Tingkat kepentingan tersebut meningkat secara proporsional dengan jumlah kata tersebut muncul dalam dokumen tetapi

di imbangi dengan frekuensi dari kata di dalam corpus. Variasi dari skema pembobotan tf-idf sering digunakan oleh mesin pencari sebagai alat utama dalam penilaian dan perangkingan relevansi dokumen dengan query yang diberikan pengguna. Term frequency merupakan frekuensi kemunculan term pada dokumen. Document frequency adalah banyaknya dokumen dimana suatu term muncul.

### Pencarian Kedekatan Dua Kata

Untuk mengukur kesamaan antara dua kata dalam penelitian ini digunakan cara pengukuran (Dao & Simpson, 2006).  
 $Sim(s, t) = 1 / jarak(s, t)$ .

$$Sim(s, t) = 1 / jarak(s, t)$$

Dimana (s) adalah kata kunci yang digunakan sedangkan (t) adalah kata penting pada dokumen. Jarak adalah jarak antara (s) dengan (t) pada ontologi.

### Implementasi

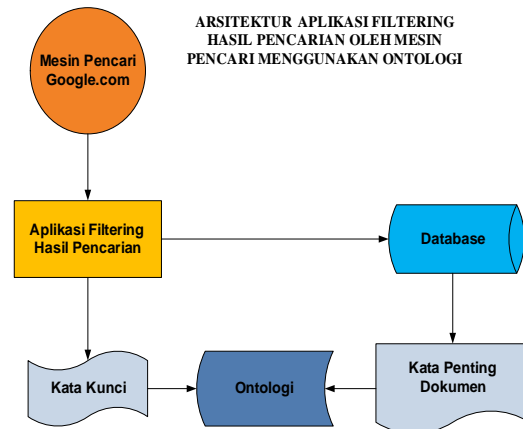
Implementasi pada aplikasi ini menggunakan bahasa pemrograman java dan database mysql.

Aplikasi dalam bentuk desktop yang digunakan untuk *filtering* hasil pencarian pada mesin pencari.

### Implementasi

Implementasi pada aplikasi ini menggunakan bahasa pemrograman java dan database mysql.

Aplikasi dalam bentuk desktop yang digunakan untuk *filtering* hasil pencarian pada mesin pencari.



Gambar 2. Arsitektur Aplikasi Hasil Pencarian Pada Mesin Pencari Menggunakan Ontologi

Pada Gambar 2 proses yang pertama kali dilakukan adalah mengambil hasil pencarian pada mesin pencari dengan cara *crawling* atau menyimpan hasil pencarian dalam bentuk *file* HTML. Kemudian *file* HTML tersebut akan diekstrak diambil informasi yang dibutuhkan meliputi judul dokumen, alamat dokumen dan isi dokumen selanjutnya disimpan kedalam *database*. Proses selanjutnya adalah mengambil kata penting pada masing-masing dokumen dengan cara memberikan bobot pada masing-masing kata dalam dokumen. Setiap kata yang mempunyai bobot paling tinggi merupakan kata penting pada dokumen tersebut. Setelah kata penting didapatkan langkah selanjutnya melakukan pencarian berdasarkan kata kunci yang digunakan pada mesin pencari setelah itu antara kata kunci dengan kata penting pada dokumen akan dicari kedekatannya pada ontologi. Apabila kata kunci dan kata penting ada dalam ontologi selanjutnya akan dihitung jarak keduanya. Proses selanjutnya dihitung menggunakan *fast heuristic* sehingga didapatkan hasil bobot pada masing-masing dokumen.

### Pengujian

#### Pengujian Dengan Satu Kata Kunci

Pengujian dengan satu kata kunci digunakan untuk melihat hasil dari relevansi

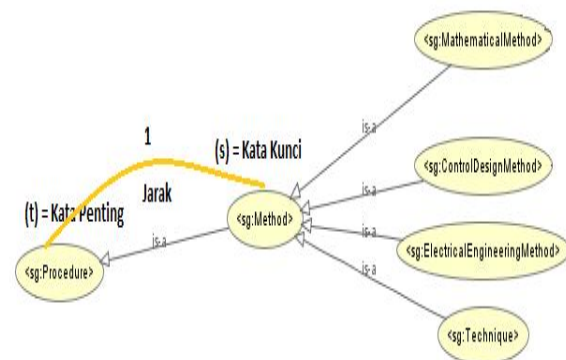
kata kunci dengan hasil pencarian. Apabila keduanya memiliki kedekatan di dalam ontologi maka jarak antar *node* akan dihitung. Pengujian ini menunjukkan bobot masing-masing dokumen. Dokumen yang memiliki bobot yang tertinggi merupakan dokumen yang relevan dengan kata kunci berdasarkan ontologi. Pada pengujian ini kata kunci yang digunakan adalah “method” dan dilakukan pengujian dengan 10 dokumen. Hasil pencarian dapat dilihat pada Tabel 1.

Tabel 1. Pengujian Dengan Satu Kata Kunci

N o	Judul Dokumen	Alamat Dokumen	Bobot
1	Home  Method	http://method. com/	0,66
2	Method (computer programmin g) - Wikipedia, the free encyclopedia	http://en.wiki pedia.org/wiki /Method_%28 computer_pro gramming%2 9	0
3	method - definition of method by the Free Online Dictionary, Thesaurus and Encyclopedi a.	www.thefreed ictionary.com/ method	0
4	Method - Wikipedia, the free encyclopedia	en.wikipedia. org/wiki/Meth od	0
5	method studios	http://www.m ethodstudios.c om/	0
6	method	methodhome. com/	0
7	What is method? definition and meaning	www.business dictionary.co m/definition/ method.html	0.66
8	Method (Java Platform SE 6)	http://docs.ora cle.com/javas e/6/docs/api/ja va/lang/reflect /Method.htm	0

9	method - Wiktionary	method – Wiktionary	0
10	method laundry	http://en.wikti onary.org/wik i/method	0

Kata kunci yang digunakan dalam pengujian menggunakan kata kunci “method”. Dari pengujian Table 1 bahwa bobot tertinggi 0,66 ada 2 dokumen sedangkan dokumen yang lain mempunyai bobot 0. Pencarian kata kunci dengan kata penting pada tiap dokumen dilakukan pada ontologi dengan mencari *node* yang ada pada ontologi. Jarak lintasan antar *node* yang ditemukan akan dihitung berapa lompatan antar *node* tersebut sehingga jarak antara kata kunci dengan kata penting dapat dihitung. Proses menghitung jarak antara kata kunci dengan kata penting dihitung berdasarkan jarak lompatan antara kata kunci dengan kata penting. Jarak kata kunci dengan kata penting akan dapat ditemukan asalkan keduanya dalam satu jalur. Apabila keduanya diluar jalur yang saling terhubung maka jarak lintasan tidak dapat dihitung jaraknya. Prinsip pencarian keduanya menggunakan jarak yang terpendek dalam lintasan. Panjang jarak antara kata kunci dengan kata penting sangat berpengaruh pada bobot yang akan dihasilkan. Semakin dekat jarak antara kata kunci dengan kata penting maka bobot yang dihasilkan akan semakin tinggi. Apabila kata kunci dan kata penting tidak ada hasilnya 0.



Gambar 3. Proses Pencarian Kata Kunci dengan Kata Penting pada Ontologi.

Dari Gambar 4.1 dapat dilihat bahwa (s) adalah kata kunci, sedangkan (t) adalah kata penting. Setiap satu lompatan antar *node* bernilai sama dengan 1. Jadi dari kata kunci *method* ke *procedure* mempunyai nilai sama dengan 1. Jarak antara kata kunci dengan kata penting sangat mempengaruhi bobot yang dihasilkan. Apabila kata kunci atau kata penting tidak ada dalam ontologi maka bobotnya adalah 0.

Tabel 2 Perhitungan Bobot

Kata Kunci	Kata Penting	
	Kata 1	Kata 2
<i>method</i>	0	1

$$\text{SumX} = \max(\text{Cell}[1,1], \text{Cell}[1,2]) = 1$$

$$\text{SumY} = \max(\text{Cell}[1,1]) = 0$$

$$\text{SumY} = \max(\text{Cell}[1,2]) = 1$$

$$\text{Matching Average} = \frac{\text{Match}(X, Y)}{|X| + |Y|}$$

$$= \frac{1 + 0 + 1}{1 + 2} = \frac{2}{3} = 0,66$$

Dimana SumX adalah nilai maksimum pada sumbu X. SumY adalah nilai maksimum pada sumbu Y. Sedangkan X dan Y adalah banyaknya baris dan kolom. Hasil rata-rata nilai kedekatan antara kata kunci dengan kata penting hasil pencarian adalah 0,66.

## KESIMPULAN DAN SARAN

Pada penelitian ini dapat disimpulkan bahwa metode yang diusulkan mampu meningkatkan relevansi hasil pencarian pada mesin pencari, bobot kedekatan kata kunci dengan kata penting yang dihasilkan dengan metode *fast Heuristic* sebesar 0,66. Semakin dekat jarak antara kata kunci dengan kata penting dokumen maka bobot yang dihasilkan akan semakin besar.

## DAFTAR PUSTAKA

- Antoniou, G., & Harmelen, F. V. (2004). *A Semantic Web Primer*. United States of America: Massachusetts Institute of Technology.
- Buckley, C., & Salton, G. (n.d.). (2010). Stop Word List 2. <http://www.lextek.com/manuals/onix/stopwords2.html>, diakses 22 April 2013
- Dao, T. N., & Simpson, T. (2006). *Measuring Similarity between sentences*. Vietnam
- Jones, B. (1994). Can punctuation help parsing. In *15 th International Conference on Computational Linguistics, Kyoto, Japan*.
- Sudeepthi, Anuradha, & Babu, S. P. (2012). A Survey on Semantic Web Search Engine. *IJCSI International Journal of Computer Science Issues*, 241-245.
- Pramudiono. (2006). *Model Pencarian pada Mesin Pencari*. Bandung.
- Wicaksana, I. W. (2006). *Ontology: Bahasa dan Tools Protege*. Depok, Jawa Barat: Universitas Gunadarma.
- Zhang, Y., Vasconcelos, W., & Sleeman, D. (2006). *OntoSearch: An Ontology Search Engine*. Scotland, UK: Department of Computing Science, University of Aberdeen, Aberdeen.
- Zouaq, A., Gagnon, M., & Ozell, B. (2009). *A SUMO-based Semantic Analysis for Knowledge Extraction*. Canada: Ecole polytechnique de Montréal, C.P. 6079, succ. Centre-ville Montreal.