Normalizing Spatial Information to Better Combine Criteria in Geographical Information Retrieval

Damien Palacio, Christian Sallaberry, Mauro Gaio

Normalizing Spatial Information to Better Combine Criteria in Geographical Information Retrieval

Damien Palacio, Christian Sallaberry, and Mauro Gaio

LIUPPA, Université de Pau, avenue de l'Université, 64000 PAU, France
damien.palacio@univ-pau.fr, christian.sallaberry@univ-pau.fr,
mauro.gaio@univ-pau.fr

Abstract. It is generally accepted that geographical information or G.I. (such as texts, maps and tables) is chiefly composed of 3 kinds of criteria : spatial, temporal and thematic criteria. The main focus of this article is spatial criteria. More specifically, we have developed a processing sequence that can extract the spatial information contained in non-structured cultural heritage texts. This processing sequence indexes spatial information, which enables information retrieval (I.R.) based on the same criteria. Our goal is to normalize heterogeneous spatial information. This normalization is carried out at the index level by grouping spatial information together and by using statistics to calculate weights of spatial areas and the pertinence of the results. Thus, we aim to develop a general IR strategy that is dedicated to spatial information, but which can be applied to temporal and thematic information as well. By generalizing this approach, homogeneous IR strategies will be able to combine spatial, temporal and thematic criteria for more efficient geographic IR methods.

1 Introduction

Our work contributes to the field of Geographic Information Retrieval (G.I.R.) as defined by [1]. The objective is to propose a complete system of GIR that will process non-structured cultural heritage texts. Therefore, we have set up an experimental platform called P.I.V. (Pyrénées Itinéraires Virtuels or Virtual Itineraries in the Pyrenees Mountains). Our works are supported by the generally accepted hypothesis that geographical information or G.I. is made up of 3 kinds of criteria : spatial, temporal and thematic criteria. One such example is : "Musical instruments in the city of Laruns in the 19th century". Indeed, this is an example of a complete unit of geographical information. To process this unit, we believe that each of its 3 components (spatial, temporal and thematic) should be treated independently, as is put forth by [2]. This can be done by making several indexes, one per component, as is advised by [3]. In this way, one can limit the search to one criterion and easily manage the indexes (e.g., to allow adding documents to the corpus). So, our approach consists in processing components independently, in order to better combine them later on. The current version of the PIV platform has 2 independent processing sequences. One is spatial and the other temporal.

However, we are now developing a way to achieve a coherent combination of indexes for each component. This means that the end-user can access more complete geographical information to carry out his research. In the field of IR, combining criteria remains a challenge according to [4]. Indeed, how can one combine criteria which have different weighting methods in information extraction (IE) and different relevancy scores in IR ? The answer is to normalize the spatial and the temporal criteria so that we come closer to time proven IR methods.

Herewith we deal with spatial criteria, while keeping in mind eventual uses for temporal or perhaps even thematic criteria.

Concerning the spatial component, existing works usually deal with web pages or short written documents (e.g. news briefs) [5, 6, 3]. On the other hand, the corpus of our study is made up of digital cultural heritage documents which have lost their page layout in the process of scanning. Contrary to web pages, our corpus is homogeneous in writing style, as almost all documents are narratives. Since these documents are travel stories, they are perfectly suitable for GIR.

In the initial version of the PIV prototype, the index is constituted by the footprint of the spatial information that has been marked [1]. During the information query stage, the key words (those that are detected as being spatial terms) are thus represented by their corresponding footprint. Then, the information is called up by matching the footprints in the index with the footprints of the query. Finally, relevancy score is computed by measuring the overlapping zone between the index footprints and the query footprints [7]. At this point, one may already perceive the difficulty in using this kind of relevancy assessment when attempting to combine it with methods used by more traditional generic IR systems.

However, with this method, each unit of spatial information in the corpus that has been detected during the marking phase is transformed independently into a directly usable index during the processing of spatial IR. Consequently, no weighting is computed for several units of spatial information within the same document unit. This is unlike the indexing techniques often suggested in the literature for traditional IR. Commonly, the weight of a term (which is used to index a document unit) is evaluated according to its frequency within the document unit and within the whole of the collection documents.

To propose a coherent solution to these problems, we suggest rearranging all the concerned spatial zones into new ones, like groups of lemmas, on to which they will be attached. The detected units of spatial information are thus

[1] Spatial data are from an IGN (French National Geographic Institute) commercial database. A footprint is a bounding polygon, a polyline or a point for datas from the IGN database and an approximated polygon for calculated datas (spatial relation)

attached to these lemmas : we call them spatial tiles. Improving spatial relevance by enhancing the context is also our objective.

Such an homogeneous spatial and temporal IR method will later allow one to coherently combine search criteria. Moreover, by combining these more precise criteria with those obtained from traditional IR systems; one gets a true GIR system.

Currently, we segment the content of digital documents into paragraphs; since the paragraph is usually considered to be the most basic segment of discourse. Here, we are referring to document units because segmentation into paragraphs is not the only way to divide a document (i.e. pages, sections). Later on it might be possible to use other book specific segments like titles, indexes and bibliographies, etc. This is further explained in [8]. Our books are mainly histories and travelogues concerning Pyrenean Cultural Heritage

In this study, we shall begin by presenting previous work, including our PIV prototype. Secondly, we will describe our proposals for creating new indexes through spatial grouping. Next, we will show the formulas we have selected to calculate spatial relevance and explain the experiments we carried out to evaluate our propositions. Finally, we will conclude by discussing our future perspectives, which deal with how to combine diverse geographical criteria.

2 Related Work

There are two types of spatial information, also called "spatial features" (SF). Simple entities correspond to named entities (e.g. "Paris" , "Vignemale peak") and complex entities are derived from simple ones (e.g. "south of Paris", "near Vignemale peak"). In our PIV prototype we consider simple entities to be absolute spatial features (ASF); whereas complex entities are qualified as relative spatial features (RSF) [9]. Temporal information is treated similarly : simple entities are of a calendar type (e.g. "spring of 1944") and complex entities are derived from these simple entities (e.g. "after summer 2003") [10]. As for thematic information, the terms are the most basic elements. Our hypothesis is that, by taking into account recent work on spatial ontologies [11], it is possible to get complex thematic entities, which are dependant on specific domains, (e.g. scientific, cultural, gastronomical,. . .) expressed in ontological form.

Regarding spatial information, of all the work which focuses on indexing simple entities (cf. table 1), only the GeoSem project [12] and PIV [9] deal with complex entities (RSFs). The Spirit prototype [13] only handles RSFs during the IR phase : it provides a selection list to the user. On the other hand, the GRID project [5] is the only one (except PIV) to offer a cartographic interface when making queries (by drawing polygons on a map). All these systems except GeoSem, are supported by geographical services (e.g. overlapping, intersection, etc.) proposed by a GIS. A variety of spatial scores based on overlaps are discussed for these prototypes (cf. [7]). Finally the STEWARD system [6] features synthetic views : it uses the frequency of ASFs to determine the reference zone associated with each document. The GeoSem project also uses statistical approaches to process the different criteria of geographical IR. Furthermore, works on spatial relevance such as [14] propositions deem that probability approaches can improve geographical IR results. This work explains that geographical information formulated within texts is often fuzzy. It is described by compressed world representations with some inaccuracy and uncertainty. So probability approaches are promoted.

Prototypes / Characteristics	PIV	STEWARD	GRID	Spirit	GéoSem
Complex entities (IE)	x				x
Complex entities (IR)	x			x	x
IR by keywords	x	x		x	x
IR by cartographic interface	x		x		
Statistics		x			x
GIS	x	x	x	x	

Table 1. Comparison of projects and prototypes used in spatial IR

Temporal aspects are seldom dealt with in the above studies. In fact, only GeoSem [12] and PIV [10] evoke it. One should note that these 2 systems both handle simple and complex temporal entities. Nevertheless, several studies discuss the temporal component in textual documents [15].
The thematic component is not presented per se as in geographical IR. It remains limited to studies on term frequency. The different prototypes (cf. table 1) generally rely on reverse lists and the TF.IDF.

Each prototype (cf. table 1) also offers a combination approach in the IR phase. Some, such as STEWARD [6] or GRID [5], prefer entry criteria then go on to apply the other criteria to a subgroup of the results (i.e. a priority of terms, then spatial information, for STEWARD). This is a sequential method and not a real combination. GeoSem [12] is the only one to combine the 3 criteria via a statistical approach (using averages). One should note that this method does not use footprints. Instead it relies on spatial, temporal or thematic thesauruses. This allows one to build less accurate indexes, but it does have the advantage of combining similarly computed relevance scores.
The main problem of the other approaches is that each relevancy calculation formula of a document unit is based on different methods corresponding respectively to spatial, temporal, or thematic criteria. The merging of results obtained from this kind of research criteria make sense only if the relevancy calculation formula is supported by similar methods. That is why in the next part we will

present a normalization spatial approach, which we will later apply to the temporal and (under certain conditions) thematic aspects.

3 Spatial Gathering to normalize

Now that we have described the features of the initial version of our prototype, which we will henceforth refer to as PIV_v1, lets move on to the second one : the PIV_v2. The innovation is to generate new indexes through spatial gathering. Figure 1 illustrates different ways to index spatial information which is relevant to a document unit. Here, document unit one (DU1) is represented by the SF vector (ASF1, ASF2, ASF3, RSF1, RSF2, RSF3, RSF4). In PIV_v2, DU1 is also represented by the tile vector (T1, T3, T4) and by the administrative county vector (C1, C2, C4, C5, C6). So, the frequency of SFs found in the tiles resulting from the administrative grid (counties) or the regular grid allows one to calculate the TF.IDF of the tiles in a document unit (just like the terms used in classical IR methods). So T3 is quoted 4 times in document unit one (DU 1), whereas C6 county is quoted only once (figure 1). These new level indexes will open new doors for digital libraries. So, a document unit can designate one or several counties. Likewise, a county can designate one or several document units.

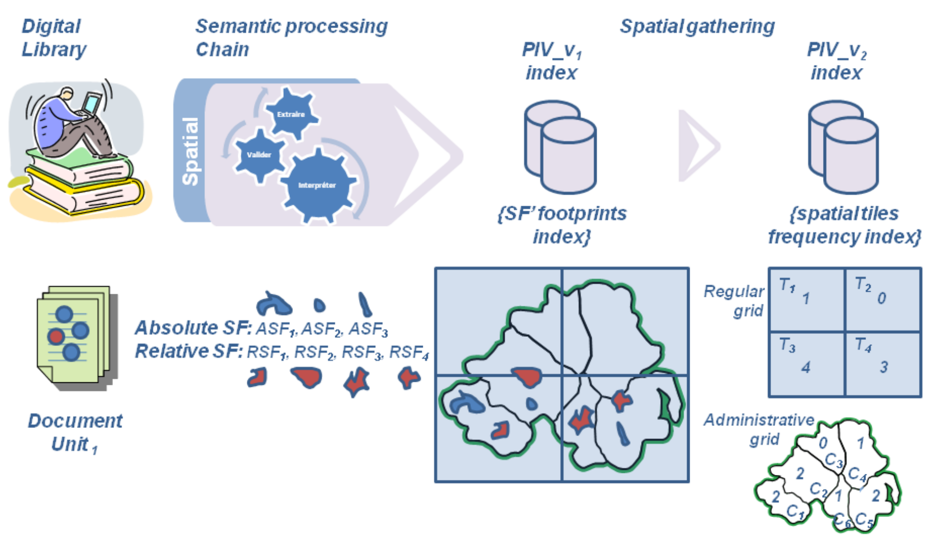

Fig. 1. SF gathering within cells of an administrative or regular grid

We propose to normalize spatial information by gathering it together. This involves dividing space into "regular tiles" (rectangular tile grids of the same size, see figure 2), or "administrative tiles" (tiles corresponding to cities, departments or districts, see figure 3). We gather nearby SFs within the same tile to emphasize their importance. This approach is comparable to better known techniques like lemmatization and term normalization.

For example, by using adminstrative segmentation (like regions) to index the SF in figure 4, we get 4 tiles (c.f. figure 6). Two of these tiles occur more frequently than the others : Aquitaine and Auvergne have a frequency of 3 (c.f. figure 6). The spatial information yielded by this document unit mainly deals with these 2 regions. Gathering together all the SFs listed in figure 4 into one and only one global zone (cf. figure 5) leads to the creation of an over scaled zone.

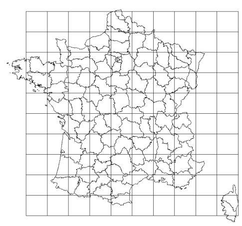

Fig. 2. Regular grid segmentation

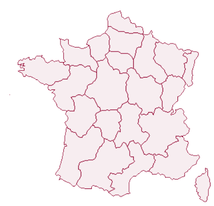

Fig. 3. Administrative segmentation

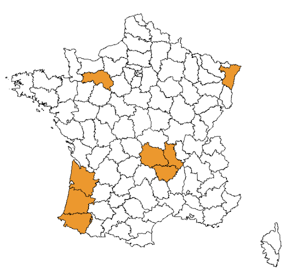

Fig. 4. SFs of a document unit

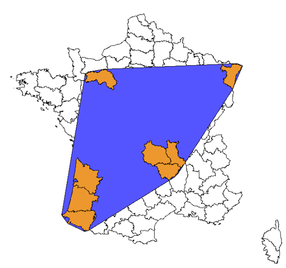

Fig. 5. Over scaled zone gathering all SFs

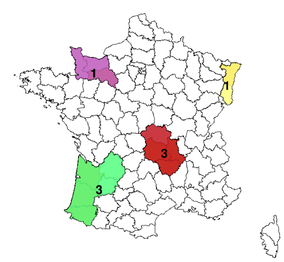

Fig. 6. Administrative segmentation with regions

Administrative segmentation (figure 3) has several advantages. First, the tiles are predefined and invariable (like regions). Secondly, they make good sense. This overlay exists for different scales according to the desired accuracy (countries, regions, departments, cities). The drawbacks which come along with this approach are of two sorts. First, it is restrictive (e.g. two neighboring departments belonging to two different regions are not gathered together). Second, this approach gathers spatial information together regardless of the difference in scale (e.g. small cities in big departments).

The regular grid (figure 2) is less restrictive than the administrative one since the size and boundaries of the tiles are adjustable. Thus, it can be defined according to the corpus. Nevertheless, each modification of the corpus may cause the grid to change as well (e.g. if a newly indexed SF is located outside the reference zone).
Moreover, it has the additional disadvantage of generating up to several million tiles if the zone is wide and the grid is narrow (e.g. overlaying the zone "France" with a grid of 1000x1000 generates one million tiles). In spite of this, it is possible to overcome this problem by using a so called "adaptable" grid, which will only handle tiles that have a high TF. For instance, by applying a 100x100 segmentation to "France", it is possible to divide again the tiles that deal with the Pyrenean zone.

In summary, spatial gathering consists in generating new indexes. For the document in figure 4 that contains 8 SFs, the new index will contain 4 tiles (cf. figure 6). A SF can overlap several tiles, like in figure 1, where RSF2 is overlapping T1 and T3. In this case, the frequency of these two tiles is incremented by 1. This is discreet indexing. It means that for a given document unit, the frequency associated with the tile is incremented by 1 if there is an intersection between the SF and the tile. We have also considered a continual indexing approach. According to the overlay ratio SF/tile, the associated frequency to the tile is incremented by a quantity of roughly between 0 and 1.

These indexes will then go on to support spatial IR. The weighting of the tiles will support results classification. We will use traditional IR formulas and carry out experiments.

As described in table 2, the TF.IDF allows one to work on tile frequency. To avoid reducing the weight of overly frequent tiles, we decided to test the TF exclusively. Since some works, such as [16], have shown that Okapi was a highly effective measure for IR; we have tested it as well.
Nevertheless, these approaches generally tend to bring spatial information to tiles through discrete indexing. That is why we propose a final weighting formula for TF, which we qualify as "continual". The latter takes the specificities of each SF inside the tile into account (e.g. the overlaying surface between the SF and the tile and the number of tiles covered by each SF). We refer to this approach as being "TF weighted". Its main purpose is to give a different weight to a peak and a city in the same tile, since these two features do not have the same size. Table 2 details these previously mentioned formulas which are used for tile weighting by PIV_v2. Furthermore, relevancy scores are computed with the inner product formula.

Tile Frequency (TF)	$W_{t,p} = TF_{t,p} = \frac{freq_{t,p}}{\sum_{i=1}^{n} freq_i}$
TF.IDF	$W_{t,p} = TF_{t,p} * IDF_t \ and \ IDF_t = log(\frac{P}{P_t})$
Okapi	$W_{t,p} = (\frac{(k_1+1)*TF_{t,p}}{(K+TF_{t,p})}) \ and \ K = k_1 * [(1-b) + \frac{b*n}{advl}]$
Weighted TF	$W_{t,p} = \sum TF_{ES,p} * \frac{Ar_{ES,t}}{Ar_t} * \frac{1}{NbT_{ES}}$
$freq_{t,p}$: tile frequency in the paragraph, n : number of tiles in the paragraph P : number of paragraphs, P_t : number of paragraph with tile t $k_1 = 1.2$, $b = 0.75$, $advl = 900$, $Ar_{ES,t}$: SF area on tile t Ar_t : tile area, NbT_{ES} : number of tiles intersected by the SF	

Table 2. Weight calculs formulas, used by PIV_v2, for a tile t and a paragraph p

4 Experiments

Here we will present the experimentation we carried out to test our different propositions for spatial normalization. Via the spatial processing sequence in PIV_v1, 10 books were indexed. This corresponds to 903 paragraphs and 10,729 SFs (9,277 ASFs and 1,452 RSFs). In order to compare our propositions to manually sorted methods, we have chosen a sample of 100 paragraphs, corresponding to 310 SF (235 ASFs and 75 RSFs). Each paragraph contains between 1 and 15 SFs. We have chosen 10 queries to be submitted to PIV_v1 and PIV_v2. Five of them have 1 ASF and the five others have 1 RSF. We have considered all the possible cases : when the ASF is completely inside the tile or when the ASF is overlapping several tiles; the same thing goes for RSFs. We have evaluated manually the relevancy of each document of the corpus, for each query. Then we have compared the results of each test with the reference evaluation.
It was not possible to use GeoCLEF corpus because PIV algorithms mainly use precise polygonal footprints and also because PIV processes are dedicated to documents written in french language.

Recall / Precision	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Average – TF	0,96	0,98	0,98	0,96	0,97	0,96	0,96	0,95	0,95	0,75
Average – TF.IDF	1	0,99	0,98	0,96	0,97	0,96	0,96	0,95	0,95	0,75
Average – Okapi	1	0,98	0,98	0,96	0,97	0,96	0,95	0,95	0,94	0,74
Average – Weighted TF	0,97	0,98	0,97	0,96	0,97	0,94	0,93	0,92	0,92	0,75
Average – PIV_v1	1	0,99	0,98	0,98	0,95	0,95	0,95	0,95	0,95	0,56

Table 3. Recall/Precision for cities segmenting

The results confirm the possibility of using classical IR formulas for spatial information. Quantitatively speaking, the results are fairly similar. Table 3 presents these results with a cities grid. We found that statistical approaches

Precision \ Recall	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Average – TF	1	0,98	0,97	0,95	0,96	0,95	0,94	0,84	0,64	0,44
Average – TF.IDF	1	0,98	0,97	0,95	0,96	0,94	0,94	0,84	0,64	0,44
Average – Okapi	1	0,98	0,97	0,95	0,96	0,94	0,93	0,83	0,63	0,44
Average – Weighted TF	1	0,98	0,97	0,96	0,95	0,94	0,94	0,83	0,63	0,43
Average – PIV_v1	1	0,99	0,98	0,98	0,95	0,95	0,95	0,95	0,95	0,56

Table 4. Recall/Precision for 100x100 grid

produce slightly better results than the PIV_v1, and they are almost identical between them. Table 4 illustrates results obtained from a regular grid of 100x100. However, this time the PIV_v1 wins out. Nevertheless, the results would be improved by using a smaller grid (e.g. 1000x1000). Testing on other samples confirm this trend, but a smaller grid multiplies the number of calculations since there are so many tiles.

From a qualitative viewpoint, we observed that the TF by itself can yield a lot of results with an identical score (score of 1,0 for 46 results of the same query). So, this approach is less satisfying than the others. Moreover, as previously mentioned for the PIV_v1, the relevancy score calculation formula does not take into account the number of SF in the overall document. Obviously, the other approaches do. Regarding weighted TF, it allows one to take into account the size of each SF (during the indexing stage) which gives more accurate calculations.

Among the different arrangements, the administrative one produces better results (cf. table 3). That is why segmenting into cities has been applied to our corpus. Indeed, our corpus is mainly composed of ASF, in particular, lots of cities.
As for segmenting into regular grids, it gives worse results than with the PIV_v1 (high recall, cf. table 4). This is especially due to choosing oversized tiles (100x100). The regular grid may end up being better in cases where fuzzy entities like our RSFs are processed.

In conclusion, the segmenting of space into tiles, combined with the use of classical IR formulas give good results. This normalization allows one to introduce an initial approximation of the context. We advise segmenting into cities and the weighted TF for a territorial corpus.
Administrative segmentation is based on an existing hierarchy (cities, departments, regions . . .). This allows one to consider information indexing and retrieval at different scales.

5 Further Work

The PIV v1 prototype was dedicated to local cultural heritage document repositories. On the other hand, the PIV_v2 approach (figure 1) aims at generalizing these spatial information results, in order to manage larger and different repositories. Nevertheless, the main purpose of our work is :

- To validate this approach through tile frequency computing.
- To associate different spatial views with document units (county, district and/or city tiles support such scale levels. Similarly, a spatial tile may be used to indicate a set of related document units).
- To propose a global geographic IR system combining spatial, temporal and thematic querying criteria (figure 7).

Consequently, we must build indexes that accumulate temporal intervals and thematic concepts according to approaches which are similar to those proposed for space (figure 7). Temporal entities that are indexed in PIV_v1 will be gathered into regular intervals of a timeline. Likewise, terms will be grouped according to the ontological concepts of the domain. Figure 7 clearly shows that PIV_v1 allows one to obtain 3 indexes of spatial features, temporal features and terms. PIV_v2 uses gathering to calculate spatial tile, time interval and concept frequency. Therefore, we have 3 independent indexes that can handle single-criteria and multi-criteria queries.

A query can be broken down into 3 criteria (thematic, spatial and temporal). Each is compared to a corresponding index (figure 7) and allows one to get a group of document units as a result. The relevancy score formulas that we propose (we have opted for the administrative segmenting, the weighted TF and the inner product for spatial information) permit one to classify these results for a given criteria. Now, we must go on to combine each group of results with those obtained by using other criteria. Merging results from this combined approach is a recurring research question nowadays [3, 13]. In fact, an IR intersection operator has been the object of experimentation for spatial and term-based queries (PIV_v1 : [17]). It guarantees accuracy, yet has a poor recall factor. Future works will deal with how to integrate spatial, temporal and thematic similarity ranking. We will also carry out experimentation on new merging algorithms by using products, maximum similarity and various linear combination functions [3].

6 Conclusion

The aim of this article is to develop a general IR strategy that is applicable to spatial, temporal and thematic information by doing gathering for each type of information. Results obtained from the PIV_v2 prototype (which is based on the spatial grouping concept) are as good as, and sometimes better than, those obtained with the PIV_v1. We note that the weighted TF approach associated with the administrative grouping of spatial information is well suited to cultural

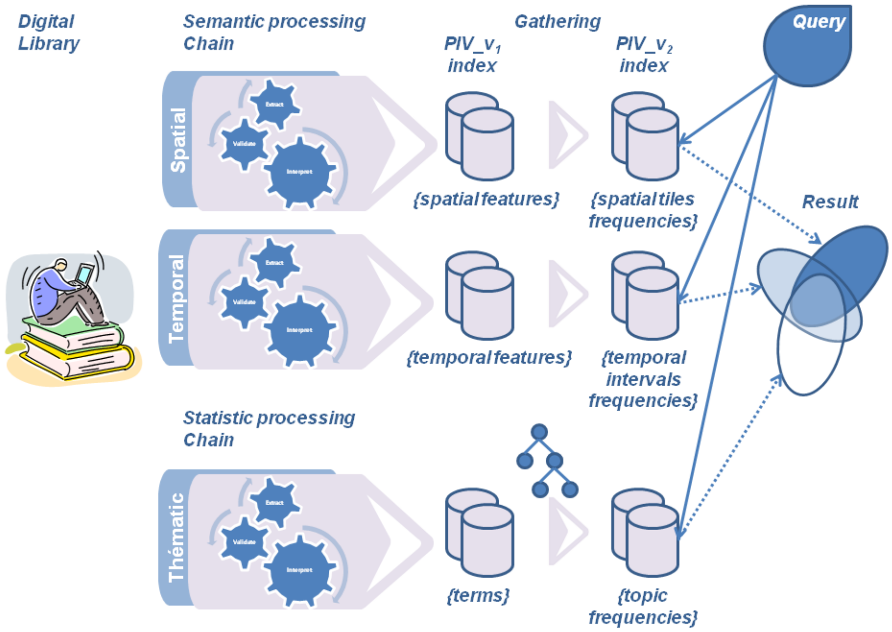

Fig. 7. The PIV spatial, temporal and thematic (textual) IE and IR processes

heritage digital libraries, especially those composed of travel stories. The concept of spatial segmentation (TF.IDF and/or weighted TF) can be adapted to the temporal and thematic index chains currently being developed for the PIV prototype. This IR work, supported by the now homogeneous weighting criteria, will facilitate the interclassification of the document units obtained in results. The main problem of current geographic IR systems comes from the fact that the weighting formulas used for space, time and theme are very different. So, they limit the possibilities of integrating results. The results presented for the PIV_v2 are a first step to resolving this problem.

We would like to further test these initial results on a larger sample of texts and queries. Then we will experiment by grouping temporal information taken from the temporal indexes already calculated by the PIV_v1. Finally, we will experiment on a combination of spatial and temporal research criteria.

References

1. Jones, C.B., Purves, R.: Gir'05 2005 acm workshop on geographical information retrieval. SIGIR Forum **40**(1) (2006) 34–37
2. Clough, P., Joho, H., Purves, R.: Judging the spatial relevance of documents for gir. In Lalmas, M., MacFarlane, A., Rüger, S.M., Tombros, A., Tsikrika, T.,

Yavlinsky, A., eds.: ECIR. Volume 3936 of Lecture Notes in Computer Science., Springer (2006) 548–552
3. Martins, B., Silva, M.J., Andrade, L.: Indexing and ranking in geo-ir systems. In: GIR '05: Proceedings of the 2005 workshop on Geographic information retrieval, New York, NY, USA, ACM (2005) 31–34
4. Visser, U.: Intelligent Information Integration for the Semantic Web. Springer-Verlag, Heidelberg (2004)
5. Valcartier: Grid - geospatial retrieval of indexed document. Technical report, R&D pour la défense Canada (2006)
6. Lieberman, M.D., Samet, H., Sankaranarayanan, J., Sperling, J.: Steward: architecture of a spatio-textual search engine. In: GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, New York, NY, USA, ACM (2007) 1–8
7. Sallaberry, C., Gaio, M., Palacio, D., Lesbegueries, J.: Extending gis functions for gir needs. In: ACM 17th CIKM, (GIR workshop), Napa Valley, CA. (2008) 1–8
8. Wu, H., Kazai, G., Taylor, M.: Book search experiments: Investigating ir methods for the indexing and retrieval of books. In Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., White, R.W., eds.: ECIR. Volume 4956 of Lecture Notes in Computer Science., Springer (2008) 234–245
9. Lesbegueries, J., Gaio, M., Loustau, P.: Geographical information access for non-structured data. In Haddad, H., ed.: SAC, ACM (2006) 83–89
10. Parc-Lacayrelle, A.L., Gaio, M., Sallaberry, C.: La composante temps dans l'information géographique textuelle. Revue Document Numérique **10**(2) (2007) 129–148
11. Velardi, P., Fabriani, P., Missikoff, M.: Using text processing techniques to automatically enrich a domain ontology. In: FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems, New York, NY, USA, ACM Press (2001) 270–284
12. Bilhaut, F., Dumoncel, F., Enjalbert, P., Hernandez, N.: Indexation sémantique et recherche d'information interactive. In: CORIA 2007. (2007) 65–76
13. Vaid, S., Jones, C.B., Joho, H., Sanderson, M.: Spatio-textual indexing for geographical search on the web. In Medeiros, C.B., Egenhofer, M.J., Bertino, E., eds.: SSTD. Volume 3633 of Lecture Notes in Computer Science., Springer (2005) 218–235
14. Larson, R.R., Frontiera, P.: Spatial ranking methods for geographic information retrieval (gir) in digital libraries. In Heery, R., Lyon, L., eds.: ECDL. Volume 3232 of Lecture Notes in Computer Science., Springer (2004) 45–56
15. Mani, I., Pustejovsky, J., Sundheim, B.: Introduction to the special issue on temporal information processing. ACM Transactions on Asian Language Information Processing (TALIP) **3**(1) (2004) 1–10
16. Savoy, J.: Morphologie et recherche d'information. Technical report, Institut interfacultaire d'informatique, Université de Neuchâtel (2002)
17. Sallaberry, C., Baziz, M., Lesbegueries, J., Gaio, M.: Towards an ie and ir system dealing with spatial information in digital libraries - evaluation case study. In Cardoso, J., Cordeiro, J., Filipe, J., eds.: ICEIS (5). (2007) 190–197