Design and usability study of an iconic user interface to ease information retrieval of medical guidelines

Nicolas Griffon,[1,2,3,4] Gaétan Kerdelhué,[1] Saliha Hamek,[5] Sylvain Hassler,[5] César Boog,[5] Jean-Baptiste Lamy,[2,3,4] Catherine Duclos,[2,3,4] Alain Venot,[2,3,4] Stéfan J Darmoni[1,2,3,4]

ABSTRACT

Background and objective Doc'CISMeF (DC) is a semantic search engine used to find resources in CISMeF-BP, a quality controlled health gateway, which gathers guidelines available on the internet in French. Visualization of Concepts in Medicine (VCM) is an iconic language that may ease information retrieval tasks. This study aimed to describe the creation and evaluation of an interface integrating VCM in DC in order to make this search engine much easier to use.

Methods Focus groups were organized to suggest ways to enhance information retrieval tasks using VCM in DC. A VCM interface was created and improved using the ergonomic evaluation approach. 20 physicians were recruited to compare the VCM interface with the non-VCM one. Each evaluator answered two different clinical scenarios in each interface. The ability and time taken to select a relevant resource were recorded and compared. A usability analysis was performed using the System Usability Scale (SUS).

Results The VCM interface contains a filter based on icons, and icons describing each resource according to focus group recommendations. Some ergonomic issues were resolved before evaluation. Use of VCM significantly increased the success of information retrieval tasks (OR=11; 95% CI 1.4 to 507). Nonetheless, it took significantly more time to find a relevant resource with VCM interface (101 vs 65 s; p=0.02). SUS revealed 'good' usability with an average score of 74/100.

Conclusions VCM was successfully implemented in DC as an option. It increased the success rate of information retrieval tasks, despite requiring slightly more time, and was well accepted by end-users.

BACKGROUND

Physicians frequently experience difficulties finding the right guideline at the right time, despite the quantity of information available.[1-3] Several quality controlled health gateways have been developed to help them. Koch[4] defined quality controlled subject gateways as internet services that apply a comprehensive set of quality measures to support systematic resource discovery. CISMeF-BP[5] (French acronym for: Catalog and Index of Online Health Resources in French—Guidelines) is one such gateway, developed at Rouen University Hospital since February 1995.[6] It now gathers 3700 guidelines or consensus development conferences, in French. To browse such a quantity of resources, a semantic search engine was created: Doc'CISMeF

(DC). This type of search engine eases the information retrieval, but has known limits. For instance, the results displayed are not all relevant to the query, requiring information searching skills that clinicians sometimes lack.[7]

Visualization of Concepts in Medicine (VCM) is a compositional iconic language created by Lamy et al.[8] VCM can represent various signs, diseases, physiological states, risks, antecedents, drug and non-drug treatments, laboratory tests, and medical follow-up procedures. It has been suggested[9] that such a language may be of interest to facilitate the use of medical search engines. The two implementations performed to date[8, 10] have both been successful: tasks were performed faster and more precisely. Icons were easier, quicker, and more convenient to read than text.[11] Using VCM in DC is a possible solution to enhance physicians' experience of information retrieval in CISMeF-BP. The aim of this work was to determine how VCM could be integrated into DC, implement it, and assess the resulting interface in terms of ergonomics and performance.

METHODS

General design of the study

The work was divided into six stages:
   First, VCM was mapped to the terminology used in DC to allow its use in the interface.
   Second, some focus groups worked on ways of facilitating the information retrieval task using VCM.
   Third, the suggestions of the focus group led to the design of a VCM interface.
   Fourth, ergonomists evaluated the VCM interface.
   Fifth, some of the issues identified by ergonomic evaluation were corrected.
   Last, the VCM interface was evaluated in comparison with the non-VCM interface.

CISMeF-BP

This catalog gathers thousands of resources described by Dublin Core meta-data[12] and manually indexed using Medical Subject Headings (MeSH).[13] Please refer to Griffon et al[14] for more information on CISMeF. As in PubMed, curators may weigh MeSH descriptors as major or minor. To find resources within CISMeF-BP, the CISMeF team previously developed DC,[15] a semantic search engine. This search engine extracts MeSH descriptors or meta-data terms from the user query and

retrieves the corresponding guidelines. The results indexed using major MeSH descriptors are displayed first. Resources with the same weight, a very common occurrence, are displayed in reverse chronological order: last published, first ranked. It is possible to perform faceted browsing[16] (figure 1C shows an example of faceted browsing). This function lists resource characteristics, which can be used to refine results. It works based on meta-data: editor, resource type, indexing term, country of publication, and date of publication. In spite of these enhancements, finding the required resource remains difficult, since too many of them are retrieved.

Visualization of concepts in medicine

VCM is a compositional language based on seven graphical components. The combination of these components allows representation of many medical concepts.[17] The seven components determine the central color, the shape, the central pictogram, the top right color, the top right pictograms, and the presence of a shadow. Each component accepts a limited number of values called 'primitives'.[8] For example (figure 2), a red icon means 'current state', and a square one means 'pathological state'. Some shapes may indicate etiology (a small chromosome

stands for 'genetic disease') or physiopathology. The central pictogram may represent the affected organ ('heart', 'lung', 'thyroid' etc.), and the top right pictogram and its color may define treatment or surveillance. The 214 primitives were organized hierarchically. Combining these primitives potentially allows generation of several million different VCM icons. In a previous study, physicians declared having spent 4–6 h learning VCM, even if some of them demonstrated good knowledge of the language after 2 h.[8]

Mappings

As documents are indexed in DC using MeSH terminology, it was first necessary to map MeSH to VCM. This mapping was performed by GK,[18] an experienced medical librarian, and completely reviewed by the three inventors of VCM to ensure maximum validity. Each MeSH descriptor within categories inside the scope of VCM (diseases category, analytical, diagnostic and therapeutic techniques and equipment category, etc.) was examined. It could be: manually mapped to VCM; or, if the MeSH concept was too fine-grained compared to VCM scope, automatically mapped to VCM, that is, it inherited VCM icons from its closest parents within relevant hierarchies (see figure 3

Figure 1   Interface of Doc' CISMeF including visualization of concepts in medicine (VCM) improvements. (A) VCM filter. Icons that share the central pictogram with the terms of the query are in the left column. The 'clinical context' icons are in the right column. All the other icons are gathered in the middle. The number between brackets is the number of resources concerned. This part of the interface is totally new compared to the old interface. (B) Result display. An iconic index allows a visual selection of retrieved resources. In the old version, VCM icons were replaced by MeSH descriptors indexing the resource (figure 4). (C) Faceted browsing. This allows refining queries using meta-data. The faceted browsing is exactly the same between the two interfaces.

Figure 2 Visualization of Concepts in Medicine iconic language. This figure shows how visualization of concepts in medicine (VCM) icons were built by summing up primitives. It is not necessary for an icon to have all the components filled (childhood pathology is built by a color, a shape, and a central pictogram only). There are only five sides for the seven components because the shadow component was not represented and color was used for the shape and the top right pictogram.
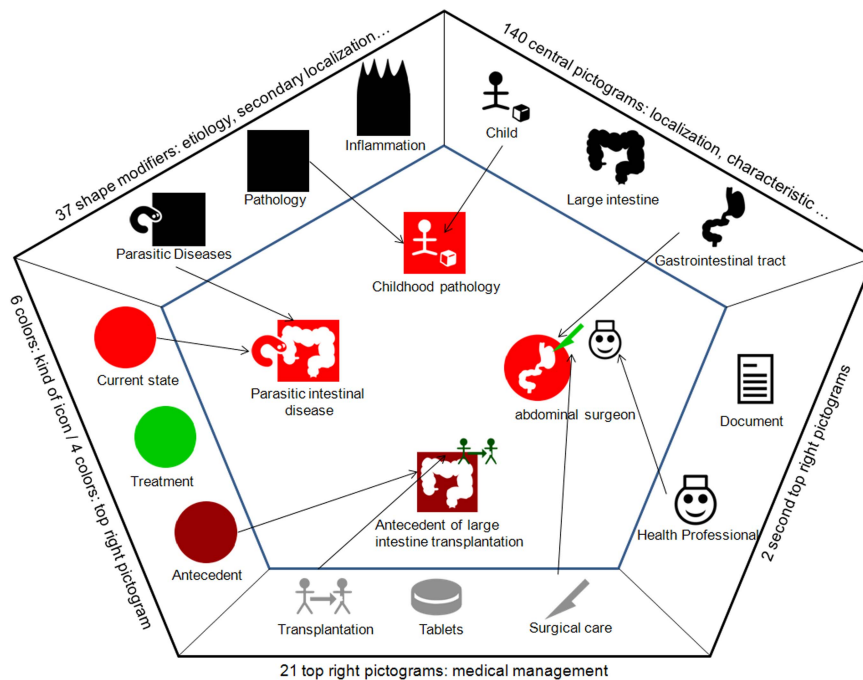
for examples). A quality assessment study of this mapping is currently under review for publication. Briefly, a kind of parallel mapping[19] was performed and showed high inter-mapper semantic agreement (>90%).

Initially, the 2011 version of MeSH was mapped to VCM. Further modifications of the MeSH thesaurus have since been taken into account. This mapping allows easy integration into any information system that works with MeSH, such as CISMeF-BP: each CISMeF-BP resource was manually indexed by MeSH descriptors and, using MeSH to VCM mapping, automatically indexed by VCM icons.

Focus group organization and design of a new CISMeF-BP interface

Focus groups, including eight general practitioners (GPs) and headed by four experienced ergonomists, were organized during two half days separated by 2 weeks in May 2009. During these 2 weeks, GPs were able to connect to a web learning platform[20] providing explanations, examples, and exercises on VCM. Only 75% of participants worked on the platform and none of them spent 4–6 h on it. Therefore, a flash training session was given

at the beginning of the focus group. This training session lasted 20 min and allowed physicians to understand the compositionality of VCM and to learn the most common primitives.

The objective of the first session of the focus group was to identify GPs' difficulties accessing information during their daily activity. For this purpose, two clinical scenarios were proposed on dyslipidemia and asthma. These clinical scenarios ended with a query to run in DC. Query results were provided to GPs printed on paper, so that GPs could underline, annotate, and highlight them. In a second time, GPs worked with the real version of the results (the web version) and made oral comments that were recorded. GPs had no access to the guidelines and were only able to access the results displayed by DC during both the paper session and the web session.

The second session was organized to determine if the iconic approach could solve the problems identified during the first session. Using paper documents, GPs were asked to use icons and to suggest context for ideal use of VCM. GPs, organized in teams of two, used iconic language to solve their difficulties, that is: How to use VCM icon to ease information retrieval? How to represent a resource with VCM icon? What information to represent? How many VCM icons? The solutions were presented to the group and argued according to the contexts of use. All oral comments from the discussions were recorded.

Based on the focus group results and the analysis of frequent queries in CISMeF-BP, the CISMeF team designed a prototype of the VCM interface.
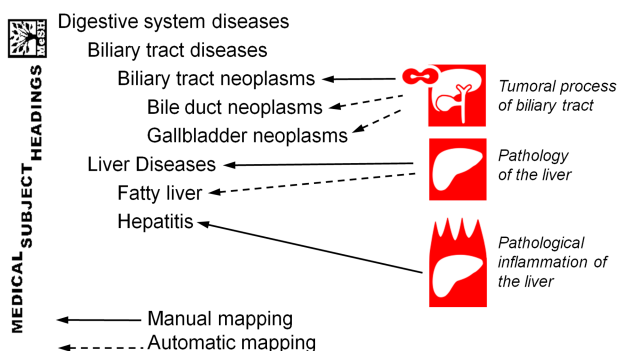
Figure 3 MeSH to visualization of concepts in medicine (VCM) mapping. When icons were not precise enough to describe MeSH descriptors, the icons were automatically inherited from their parents.

Ergonomic evaluation of the new DC interface

This prototype was evaluated by four ergonomists using the criteria described by Bastien and Scapin.[21] Each ergonomic issue encountered was reported with: the ergonomic criteria concerned, the problem description, and the gravity of the problem in terms of interface usability. When possible, recommendations were provided to improve the prototype. The CISMeF team made some of the enhancements proposed by the ergonomic study, focusing on major problems and 'easy-to-fix' problems.

Usability study design

The evaluation used four clinical scenarios (asthma, dyslipidemia, pneumonia, and urinary tract infection). They were composed of a brief clinical case concluded by a clinical question and a query to run in DC (see supplementary appendix for clinical scenarios). Evaluators could not change the queries. They were asked to select as many resources as they wished, based on the results displayed in CISMeF-BP alone. To limit the duration of the study for participants, they were not supposed to read the whole resource, a previous study having shown that approximately 90% of GPs found the relevant information in the resource once they had located it.[22]

DC queries voluntarily lacked precision in order to increase the information retrieval task difficulty by retrieving many results. Furthermore, internal analysis of logs showed that the most frequent queries run by general users of the CISMeF website only included one disease name. A medical resident (NG) and a cardiologist (PM, see acknowledgments) independently classified the resources retrieved in clinical scenarios as relevant or irrelevant; disagreements were discussed and resolved. Contrary to evaluators, they had access to the totality of resources and were therefore certain of the presence or absence of the answer. Twenty physicians were recruited for this study: 10 from Rouen University Hospital (mainly from the emergency department) and 10 GPs from the French Society of General Medicine. Each of them evaluated the four clinical scenarios: two with the VCM interface, and two with the non-VCM interface. The order of performance of clinical scenarios and the order of use of interface version were controlled in the experimental design (see table 1).

Evaluators were supposed to have used the learning platform[20] before evaluation, but log analysis showed that only 30–40% of evaluators did so. In order to limit this bias against VCM, evaluators performed the same flash training as GP focus groups. Moreover, one clinical training scenario was created to allow evaluators to discover the VCM interface of DC. The evaluation started when the evaluator felt confident with this interface. All evaluators provided written informed consent prior to any study procedures. Ethical committee approval for this type of study is not required.

Evaluation took place in a quiet dedicated room in September 2011, using computers provided by the evaluation team (with recording software, webcam, and microphone included). During the whole process, evaluators were asked to think aloud and were recorded (computer screen, audio/video of the evaluator)[23] for satisfaction analysis and time measurement (CB and SyH). After the evaluation, the evaluators were asked to complete a System Usability Scale (SUS)[24 25] questionnaire and were interviewed for further satisfaction evaluation. The SUS was composed of 10 statements scored on a scale of five modalities for strength of agreement. The final score ranged from 0 to 100, usability and user satisfaction increasing with this score (see supplementary appendix for score computation, interpretation, and question list).

Performance of the information retrieval task was assessed using two main measures:

▶ The success of each information retrieval task (whether the evaluator selected a resource classified as relevant). This was the main criterion, answering the main question: 'Are there any differences between the two interfaces in terms of success rate of information retrieval tasks?'

▶ The time spent on the information retrieval task (hereafter the 'response time'), for evaluations in which the first selected resource was pertinent. Time was measured from the launch of the query to selection of the first resource. It represents only a fraction of the time required for a user to find the answer to his/her question, but it is the only time which is measurable in this study. The other fraction of time needed to find an answer (ie, reading the document to find the relevant information) was hypothesized to be equal between groups. Response time allows us to answer the following question: 'Are there any differences between the two interfaces in terms of time to find a relevant resource?'

Statistical analysis

Concordance between NG and PM was evaluated using the κ statistic.[26] Success rates according to VCM integration were compared using Fisher's exact test. Success rates according to other controlled factors—(i) clinical scenarios (asthma vs pneumonia etc.), (ii) order of clinical scenarios (first two clinical scenarios vs last two clinical scenarios), (iii) order of interface use (VCM interface first vs VCM interface last)—were also compared.

Response time was log-transformed and compared using analysis of variance. Homoscedasticity was tested using the Bartlett test. Controlled factors were added to the model. Kaplan–Meyer survival curves were drawn to represent success rate according to time. Curves were compared using the log-rank test.

All tests were two-sided. Statistical analyses were performed using R V.2.14.2 with the 'survival' package.

RESULTS
Mappings

A total of 1830 MeSH descriptors were manually mapped and 8953 were automatically mapped to VCM. One or more VCM icon was assigned to 10 783 MeSH descriptors out of 26 142 (41.2%); 1070 VCM icons were mapped to one or more MeSH descriptors.

Focus group results and design of a new CISMeF-BP interface

When using DC, physicians encountered three main difficulties:

▶ The texts associated with the title of each resource were found to be too long, with too many keywords and therefore difficult to read.

Table 1 Experimental design of the usability study

| | Clinical scenario | | | |
	Dyslipidemia	Asthma	Urinary tract infection	Pneumonia
Interface 5 evaluators	VCM (+)	VCM (+)	VCM (−)	VCM (−)
Order	1	2	3	4
Interface	VCM (+)	VCM (+)	VCM (−)	VCM (−)
Order	3	4	1	2
Interface	VCM (−)	VCM (−)	VCM (+)	VCM (+)
Order	1	2	3	4
Interface	VCM (−)	VCM (−)	VCM (+)	VCM (+)
Order	3	4	1	2

Five evaluators resolved first the dyslipidemia clinical scenario using VCM interface, second the asthma clinical scenario with the VCM interface, third the urinary tract infection scenario with the non-VCM interface etc.
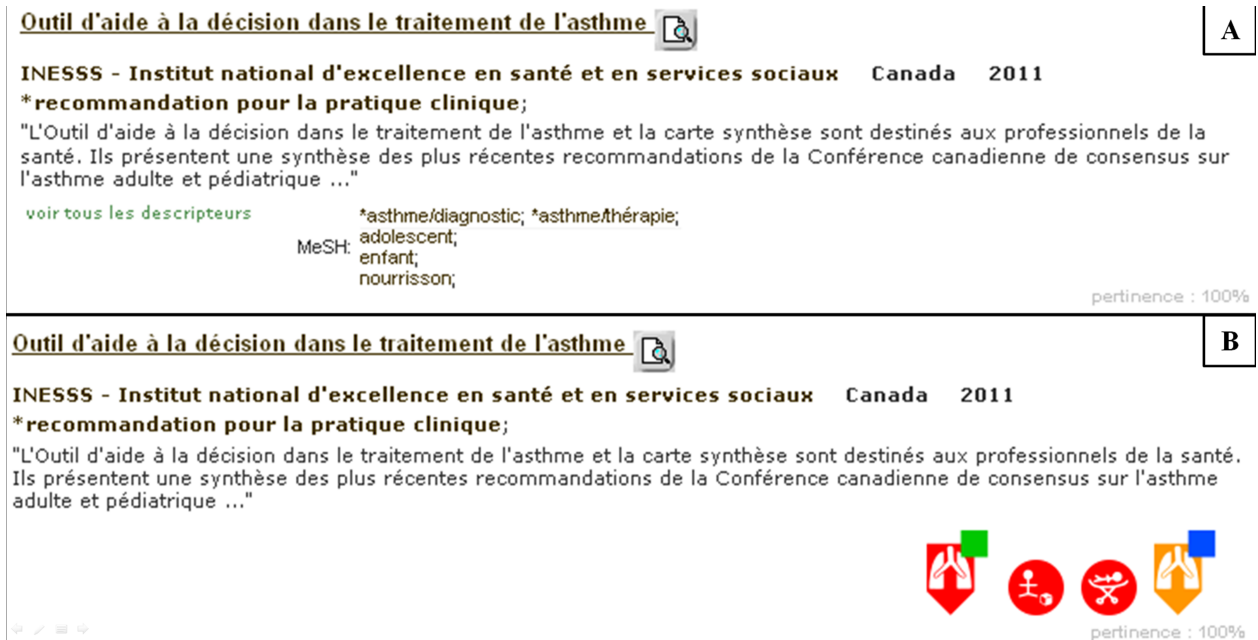VCM, Visualization of Concepts in Medicine.

Figure 4 Integration of visualization of concepts in medicine (VCM) icons in displayed results of Doc'CISMeF. (A) Doc'CISMeF without VCM icons; MeSH indexing is provided. (B) Doc'CISMeF with VCM icons; MeSH indexing is not provided.

▶ Physicians noticed the presence of resources, which were not clearly related to their queries: some results of DC were irrelevant.

▶ The relatively high number of retrieved resources was difficult to browse. A preselection filter would probably limit this difficulty.

In order to limit these difficulties, VCM was integrated into the CISMeF-BP interface in two ways:

▶ First, to reduce the amount of information in the results, MeSH descriptors were discarded and replaced by VCM icons mapped to major MeSH descriptors used to index resources (see figures 1B and 4). This allowed a visual selection of retrieved resources by end-users who might simply exclude resources not clearly related to the query using VCM icons. These VCM icons were added for each resource retrieved by DC.

▶ Second, to optimize information retrieval, a VCM based filter was added (see figure 1A). This filter gathered all the VCM icons present in any retrieved resource. It was separated in three columns, each being sorted by icon frequency. The left column gathered all the icons sharing the central pictogram with the query term. The right column allowed filtering on circle icons (the circle shape meaning 'physiological state') corresponding to clinical context (infancy, pregnancy, etc.). The middle column gathered all the other icons.

These integrations allowed the end-user to benefit from VCM icons when searching guidelines in CISMeF-BP.[27]

Ergonomic evaluation of the new CISMeF-BP interface

Ergonomic evaluation revealed 31 issues: 16 of minor and 15 of major gravity. Of these, seven were corrected by the CISMeF team and had therefore disappeared for the evaluation presented here: icon superimposition, time of appearance of tooltips, conflicts between VCM filter and faceted browsing (n=2), unusual link in the resource iconic index, modification of the relevance score of resources with the same query using iconic index link, and deselection of every check box with deselection of one of them.

Five issues depended only on VCM and therefore did not fall within the competence of CISMeF: the 'age range' icons were not precisely defined, some icons were very similar, icon labels were confusing (n=2), and interpretation of icon for treatment was not clear (see, eg, icons of figure 1B: did the treatment induce a decrease of cardiac function or did it treat it?).

Concerning the 19 problems that were not fixed, and according to ergonomists:

▶ Thirteen were of minor gravity (see supplementary appendix for complete list).

▶ Six were of major gravity: icons classification in the filter was not intuitive, the filter contained too many icons, there was no 'reset filter' button, position of check boxes was not consistent, the relevance score, which concerns the resource may be interpreted as the relevance score of the VCM icon (see figure 4), and there is apparent discordance between icon label and icon tooltip.

Performance evaluation results

PM and NG showed 'excellent' agreement according to Landis and Koch,[28] with $\kappa=0.82$. The 20 evaluators were included and each responded to the four clinical scenarios. They failed to select a relevant resource in 10 cases (12.5%; 95% CI 7% to 22%). There were differences according to interface: only one failed case with VCM (2.5%; 95% CI 0% to 13%) versus nine failed cases without VCM (22.5%; 95% CI 12% to 38%)—case specific results are available in the supplementary appendix. This difference was significant (p=0.01). Other tested factors were not significantly associated with the success of the information retrieval task (see table 2).

For 14 evaluations, the first resource selected was not relevant. These evaluations were then excluded from the time analysis. The mean response time per clinical scenario was 85 s (95% CI 69 to 100). This time varied significantly according to interface. The other controlled factors were not

Table 2 Associated factors with risk of failure (univariate analysis; n=80)

	OR	95% CI	p Value
Interface			
With VCM	Ref.		0.01
Without VCM	11.0	(1.4 to 507)	
Order of VCM use			
First	Ref.		1
Second	1.0	(0.2 to 4.8)	
Order of clinical scenario			
First two	Ref.		0.7
Last two	0.6	(0.1 to 2.9)	
Clinical scenario			
Asthma	Ref.		0.37
Dyslipidemia	3	(0.5 to 18)	
Urinary tract infection	1.0	(0.1 to 7.9)	
Pneumonia	0.5	(0.04 to 5.7)	

'Ref.' means that this is the category used for comparison.
Evaluators had an 11-fold likelihood of finding an irrelevant resource when using non-VCM interface and this is significant (p=0.01).
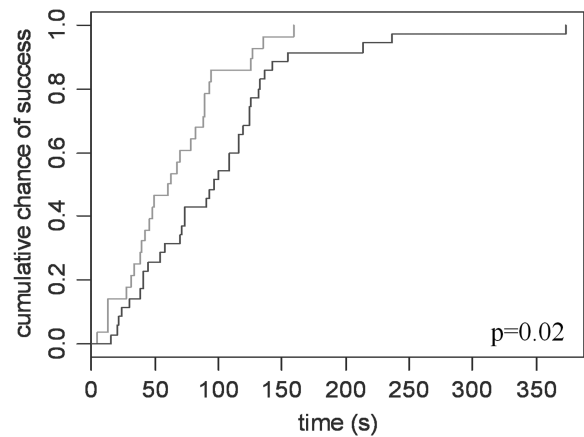VCM, Visualization of Concepts in Medicine.

Figure 5 Cumulative chance of success according to time and visualization of concepts in medicine (VCM) use. Time was considered at selection of the first resource. Failed evaluations were discarded. Grey: without VCM (n=27); black: with VCM (n=36).

significantly associated with response time (see table 3). Multivariate analysis showed similar results. The homoscedasticity hypothesis was not rejected (p>0.05; Bartlett test) for any model.

Figure 5 represents the cumulative rate of success with VCM (black curve) and without VCM (gray curve) according to time. The analysis revealed that evaluators found relevant resources significantly quicker without VCM than with VCM (p=0.02).

Satisfaction results

All the evaluators completed the SUS questionnaire. The satisfaction score revealed by the mean SUS value was good according to the Bangor et al[25] scale: 74/100 (see supplementary appendix for detailed results and interpretation). Audio files and interviews allowed investigators to discover comments on VCM integration. At first, this revealed several strengths: (1) most evaluators appreciated the relevancy of being able to expand their queries using the VCM filter; (2) evaluators felt that such a filter could speed up their information retrieval tasks; and (3) the right-hand column was understood quickly, thus allowing a quick and easy way to expand queries. Nevertheless, this qualitative part of the study also revealed that not all users understood or found the first two columns of the VCM filter easy to use: too many icons were found to be very similar.

DISCUSSION AND CONCLUSION

This study reveals that VCM interface significantly improves user results, providing a lower failure rate. Nonetheless, for those finding a relevant resource this took significantly more time. The VCM interface was well accepted by users with a mean SUS score of 74, which is good (see supplementary appendix).

Nelson et al[29] suggested that access to semantic knowledge was more efficient through graphical clues. Wiedenbeck[30] listed some of the supposed advantages of icons over text: icons are better than words for representing subtle visual and spatial concepts, icon speed search, icons lead to immediate recognition, icons lead to better recall, icons reduce the necessity of reading, and icons make interfaces more international. This picture superiority effect has been well debated and it is still unclear if the use of pictures enhances or impedes the usability of an interface.[31] This study will not end the discussion, but user satisfaction is a strong indicator that iconic languages is of interest in the human–computer interface and therefore require further research.

Results interpretation

The selection of an a priori relevant resource took more time with the VCM interface. As time is frequently reported as the main limiting factor for physicians researching information on the internet,[1–3] it could counteract the benefits of such an interface. Nevertheless, the VCM interface increased the relevance of resource selection. The time saved by selecting a relevant resource was not measured in this study, but it is likely sizeable since reading an irrelevant resource and having to look for another one adds time to the overall search process. It is not possible to say, with this study, if the time saved by the higher probability of finding a relevant resource compensated for the

Table 3 Response time according to controlled factors (n=63; ANOVA)

	Response time (in seconds)		
	Mean	95% CI	p Value*
Interface			0.04
With VCM	99	(75 to 123)	
Without VCM	66	(50 to 82)	
Clinical scenario			0.3
Asthma	81	(56 to 105)	
Dyslipidemia	110	(68 to 153)	
Urinary tract infection	93	(46 to 141)	
Pneumonia	66	(47 to 85)	
Order of VCM use			0.4
First	80	(63 to 97)	
Second	91	(63 to 120)	
Order of clinical scenario			0.06
First two	91	(74 to 108)	
Last two	79	(53 to 105)	

VCM, Visualization of Concepts in Medicine.

time lost selecting the resource. However, there are some indicators that there may be less time lost using VCM, thus increasing the interest of the VCM solution:

▶ The two main explanations for the increase in time in information retrieval tasks are: (1) evaluators may spend time filtering resources using the VCM filter (figure 1A), hence leading to a more relevant selection; and (2) evaluators may also take a few seconds to check the relevance of a resource with its VCM icons (figure 4). These two techniques may be combined in any proportion. The qualitative analysis (video review) revealed that users spent a lot of time on the VCM filter, suggesting that the first explanation is the most relevant. Regular use of the interface and/or simplification of the VCM filter may decrease response time.

▶ A similar situation arises with VCM. Even if it was quite an easy language, it took 4–6 h to learn.[8] Not all the evaluators had such time and only received a single crash training session just before the evaluation. Therefore, the evaluators' knowledge of VCM was limited, which might explain the length of time taken to select a resource using the VCM interface. The integration of VCM icons in many health-dedicated tools (patient records, drug monographs, guidelines etc.)[9] may enhance physicians' understanding of these icons, and similarly, the efficiency of the VCM interface. VCM icons involved red and green color. This can seriously limit the use of VCM in a medical interface as color blindness prevalence can reach about 8% in white male Europeans.[32] Nevertheless, color information is often redundant with position information: 91.4% of shape modifiers are red (never green) and 90.4% of top right pictograms are green (never red).

The evaluators did not appreciate the left and middle columns of the VCM filter where too many icons were presented, which was one of the uncorrected ergonomic issues. On the contrary, many of them appreciated the right-hand column. Unfortunately, this may be the result of the study design. The queries were very short, on purpose, to artificially increase the difficulty of information retrieval tasks. Therefore, they never included information on patient context—pregnancy, old age etc.—that a real life physician might have added directly to the query.

Limitations

Many ergonomic issues revealed by the ergonomic study were not corrected in the final version (n=24/31). In fact, following all the ergonomists' recommendations would have required unavailable information (eg, a more precise indexing of CISMeF resources), and huge changes in DC that were not realistic according to the CISMeF information system (eg, filtering with icons' primitive instead of pre-coordinated icons). Some of these changes would have confronted the user to the 200 primitives, which would certainly not have enhanced the usability. Nevertheless, it is an important limit to an ergonomic study. However, this is also a strength of VCM: even with numerous ergonomic issues, VCM integration shows good usability.

Working with only four clinical scenarios raises concerns about the applicability of the results to any query run in any search engine. Combined with the limited number of evaluators, this also induces a lack of power. It is likely that with greater power some other factors would have appeared to be significant. The difficulty level of each clinical scenario was not assessed. It would have been interesting to see if rough VCM language enhances relevance selection similarly for 'easy' clinical scenarios and for more complex ones.

Evaluators' characteristics were poorly documented in this study. GPs and emergency medicine physicians were studied because we believe that they are more likely to be confronted with heterogeneous clinical situations in which a clinical guideline search engine may be useful. Other specialties were not studied and it is not sure that the same results would have been observed. No other confounders were studied (age, informatics skills, information retrieval experience, etc.), but as each evaluator is his own control (using both interface), this should not introduce any bias.

Some authors have suggested that the 'thinking aloud' method introduces bias in time measurement.[33] It is possible that verbalizing icons, which roughly corresponds to transforming into textual information, may limit the aimed cognitive interest of VCM: embodying an idea with expressive icons. Nevertheless, the eventuality of a differential bias in time measurement is unlikely considering the low speech flow observed on video, for either VCM or non-VCM evaluation.

VCM was integrated into the interface via a filter and iconic indexing of resources. This implies many choices: How to sort icons in the filter? How to sort indexing icons? In terms of human interface, where to put the filter? Where to put indexing icons? These choices were based on focus group and ergonomic study, however only one possible integration of VCM was tested in the current study. Other integrations might perform better than this one. Reducing the VCM filter to a faceted search on the 'clinical context' (the third column of the filter) may improve the understandability of the interface and reduce time 'lost' on the current filter. Nevertheless, evaluators expressed an interest in such an iconic language, justifying further research on this topic.

Conclusion

Because of the better results selected using the VCM interface, the good qualitative evaluation of VCM integration in DC, the good usability revealed by SUS, and physician enthusiasm for VCM iconic language, the iconic search has been extended to each version of this search engine: (1) the generic search engine,[27] but also the more specific ones: the search engine devoted to (2) students, and (3) patients. These versions can be optionally activated on the CISMeF home page. Moreover, VCM iconic language and more specifically the VCM iconic representation of each MeSH descriptor are displayed in the CISMeF terminology portals.[34 35]

In conclusion, the results of our study demonstrate that the integration of icons into a medical search engine, despite requiring slightly more time, improves search success and is well accepted by end-users.

Provenance and peer review Not commissioned; externally peer reviewed.

VCM language is protected by an international patent submitted by Paris 13 University.

REFERENCES

1. Coumou HC, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. *J Med Libr Assoc* 2006;94:55–60.
2. Graber MA, Randles BD, Ely JW, *et al*. Answering clinical questions in the ED. *Am J Emerg Med* 2008;26:144–7.
3. Bernard E, Arnould M, Saint-Lary O, *et al*. Internet use for information seeking in clinical practice: a cross-sectional survey among French general practitioners. *Int J Med Inform* 2012;81:493–9.
4. Koch T. Quality-controlled subject gateways: definitions, typologies, empirical overview, subject gateways. *Online Inf Rev* 2000;24:24–34.
5. CISMeF. http://www.cismef.org (accessed Dec 2012).
6. Darmoni SJ, Leroy JP, Baudic F, *et al*. CISMeF: a structured health resource guide. *Methods Inf Med* 2000;39:30–5.
7. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. *Health Info Libr J* 2007;24:78–94.
8. Lamy JB, Duclos C, Bar-Hen A, *et al*. An iconic language for the graphical representation of medical concepts. *BMC Med Inform Decis Mak* 2008;8:16.
9. Lamy JB, Duclos C, Hamek S, *et al*. Towards iconic language for patient records, drug monographs, guidelines and medical search engines. *Stud Health Technol Inform* 2010;160:156–60.
10. Lamy JB, Venot A, Bar-Hen A, *et al*. Design of a graphical and interactive interface for facilitating access to drug contraindications, cautions for use, interactions and adverse effects. *BMC Med Inform Decis Mak* 2008;8:21.
11. Horton W. *The icon book: Visual Symbols for Computer Systems and Documentation*. New York: John Wiley and Sons, 1994.
12. Darmoni SJ, Thirion B, Leroy JP, *et al*. The use of Dublin Core metadata in a structured health resource guide on the internet. *Bull Med Libr Assoc* 2001;89:297–301.
13. Nelson SJ, Johnson WD, Humphreys BL. Relationships in Medical Subject Heading. In: Bean CA, Green R, eds. *Relationships in the organization of knowledge*. Dordrecht: Kluwer Academic Publishers, 2001:171–84.
14. Griffon N, Soualmia LF, Névéol A, *et al*. Evaluation of multi-terminology super-concepts for information retrieval. *Stud Health Technol Inform* 2011;169:492–6.
15. Darmoni SJ, Thirion B, Leroy JP, *et al*. Doc' CISMEF: a search tool based on 'encapsulated' MeSH thesaurus. *Stud Health Technol Inform* 2001;84:314–18.
16. Tunkelang D. *Faceted Search. Synthesis Lectures on Information Concepts, Retrieval, and Services*. Morgan and Claypool Publishers, 2009:1–80.
17. VCM: An iconic language for the Visualization of Concepts in Medicine. http://vcm.univ-paris13.fr Last accessed April 2013.
18. Kerdelhué G, Lamy JB, Venot A, *et al*. An iconic language for the 'CISMeFBonnespratiques' website [poster]. 12th European Association for Health Information and Libraries conference, Lisbon 2010 Portugal.
19. Wieteck P. Furthering the development of standardized nursing terminology through an ENP-ICNP cross-mapping. *Int Nurs Rev* 2008;55:296–304.
20. Training platform for VCM. URL: http://vcm.univ-paris13.fr/content/vcm-tutorial (accessed Dec 2012).
21. Bastien JMC, Scapin DL. Ergonomic criteria for the evaluation of human-computer interfaces. Rocquencourt: INRIA Technical Report n°156; 1993.
22. Baert-Quibel L. Évaluation de l'outil de recherche Doc'Cismef pour la diffusion des recommandations professionnelles sur l'internet: Une étude quasi-randomisée auprès des médecins généralistes. MD thesis. Rouen: Rouen University, 2004:112 p.
23. Cimino JJ, Patel VL, Kushniruk AW. Studying the human-computer-terminology interface. *J Am Med Inform Assoc* 2001;8:163–73.
24. Brooke J. SUS. a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, *et al*. eds. *Usability Evaluation in Industry*. London: Taylor and Francis, 1996:189–94.
25. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud* 2009;4:114–23.
26. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
27. CISMeF. CISMeF BP. [http://cisdev.chu-rouen.fr/servlets/CISMeFBPvcm]. Login and password are available on demand.
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
29. Nelson DL, Reed VS, Walling JR. Pictorial superiority effect. *J Exp Psychol Hum Learn* 1976;2:523–8.
30. Wiedenbeck S. The use of icons and labels in an end user application program: an empirical study of learning and retention. *Behav Inf Technol* 1999;18:68–82.
31. Oates JM, Reder LM. Memory for pictures: sometimes a picture is not worth a single word. In: Benjamin AS.ed *Successful remembering and successful forgetting*. London: Taylor and Francis, 2011:447–61.
32. Birch J. Worldwide prevalence of red-green color deficiency. *J Opt Soc Am A Opt Image Sci Vis* 2012;29:313–20.
33. Bakhshi-Raiez F, de Keizer NF, Cornet R, *et al*. A usability evaluation of a SNOMED CT based compositional interface terminology for intensive care. *Int J Med Inform* 2012;81:351–62.
34. Grosjean J, Merabti T, Dahamna B, *et al*. Health multi-terminology portal: a semantic added-value for patient safety. *Stud Health Technol Inform* 2011;166:129–38.
35. Portail terminologique de santé. URL: http://pts.chu-rouen.fr (accessed Dec 2012).