



HAL
open science

Making sense of technical documentary heritage. The contribution of terminology and domain ontology

Caroline Djambian

► **To cite this version:**

Caroline Djambian. Making sense of technical documentary heritage. The contribution of terminology and domain ontology. AIDAinformazioni: Rivista di Scienze dell'informazione, 2019. hal-04604710

HAL Id: hal-04604710

<https://hal.univ-grenoble-alpes.fr/hal-04604710>

Submitted on 7 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dare senso ai patrimoni documentali tecnici: il contributo della terminologia e dell'ontologia di dominio

CAROLINE DJAMBIAN*

ABSTRACT: Organizations have experienced for 40 years mass computerization, advent of the Internet and dematerialisation of documents, moving from document management towards managing the overflow of information and nowadays data. Meanwhile, their functioning evolves from top-down to transversal approaches, and their contexts require an operational mobilization of information and knowledge. The response of the organizations is generally sought in the ICT tools, whose accumulation makes the problem more complex. We present here a specific case illustrating the need to enhance the organisations' knowledge heritage. We explain why upstream contextualization work is essential, in order to situate the operating mode of the SI in a larger structural problem. In this context, the core business of the company appears as the basis of any reflection. Their terminology is the key we use to improving better management of the documentary heritage and knowledge they transmit, through a knowledge database based on a domain ontology.

Keywords: Technical document, Knowledge management, Knowledge base, Ontology, Terminology.

1. Introduzione

Il presente lavoro di ricerca è stato condotto nell'arco di tre anni presso la Divisione di Ingegneria Nucleare (DIN) di Électricité de France (EDF), multinazionale francese dell'energia. La DIN conta 3.500 dipendenti ed è responsabile della progettazione delle centrali nucleari attualmente in funzione, dei nuovi impianti in costruzione (in particolare, il progetto EPR) e delle centrali nucleari in corso di smantellamento sul territorio nazionale francese.

All'arrivo in azienda, il management non ha saputo esprimere chiaramente il suo bisogno, concentrandosi solo su problematiche di incapacità a gestire il volume di informazioni, l'accesso alle conoscenze aziendali e la perdita di queste conoscenze essenziali per l'azienda al momento del pensionamento dei suoi ingegneri. In realtà, tante organizzazioni (private o pubbliche) si trovano ad affrontare questo tipo di problemi: la loro attività poggia sulla produzione intensiva di documentazione tecnica. L'aumento esponenziale del volume di informazioni tecniche da gestire va di pari passo con l'esplosione delle tecnologie digitali, così che il patrimonio documentale costituito nel corso della loro storia si è spesso accumulato senza che le aziende abbiano avuto il tempo di adattarsi al crescente ritmo degli sviluppi delle tecnologie dell'informazione. Il patrimonio di conoscenze che viene quindi incessantemente prodotto, e che costituisce il fulcro dell'attività tecnica delle imprese, ha un volume crescente, è diffuso ed eterogeneo¹. Parallelamente, problematiche trasversali e strategiche richiedono oggi alle organizzazioni di poter mobilitare queste conoscenze in modo concreto e operativo. Tuttavia la questione ha assunto spesso dimensioni tali da non sapere più come affrontarla.

* Université Grenoble Alpes – France, GRESEC (Groupe de REcherche Sur les Enjeux de la Communication)
caroline.djambian@univ-grenoble-alpes.fr.

¹ F. GANDON, R. DIENG-KUNTZ, *Ontologie pour un système multi-agent dédié à une mémoire d'entreprise*, in *Ingénierie des connaissances*, L'Harmattan, Paris 2005.

Infatti, i tentativi delle organizzazioni di risolvere questi problemi a volte non hanno successo perché esaminando il contesto appare in maniera evidente che la questione va ben oltre gli aspetti tecnici e riguarda l'organizzazione nel suo complesso. Troppo spesso si crede di poter risolvere questi problemi con le sole soluzioni tecniche, che vengono così accumulate peggiorando le cose. Ma in questi casi la soluzione va piuttosto ricercata al livello globale dell'azienda, che in un dato momento non riesce più a conciliare la documentazione prodotta e conservata e il bisogno informativo degli utenti.

Nell'ambito della DIN di EDF, le osservazioni effettuate ci hanno quindi imposto di prendere in considerazione esigenze che comportavano un'ampia riflessione sia in termini di valorizzazione del patrimonio documentale, che di capitalizzazione della conoscenza dei *core business* dell'azienda. Al fine di affrontare gradualmente la situazione da risolvere, il lavoro è stato basato sull'analisi del problema nel suo complesso, attraverso il Sistema Informativo e poi attraverso quello dei *core business* e del contesto trasversale. In un secondo momento, allo scopo di evolvere verso un'organizzazione più coerente del patrimonio documentario e delle conoscenze in esso contenute, è stata dapprima identificata un'area particolarmente sensibile, gli *incidenti nucleari gravi*, e si è iniziato a riflettere sulla costruzione di una base di conoscenza *core business* dedicata a questo dominio. Questo ha portato a mettere in discussione i metodi scientifici di gestione e organizzazione della conoscenza e in particolare, quelli della costruzione di terminologie e ontologie specializzate. Viene quindi di seguito presentato l'approccio termino-ontologico scelto e il metodo adottato per la costruzione della base di conoscenza "incidenti gravi".

2. Analisi preliminare: descrizione del problema

2.1. Andare oltre il Sistema Informativo

Sia per la gestione della documentazione che delle conoscenze, se si vogliono risolvere in modo duraturo le questioni complesse che si sono create nel tempo all'interno delle organizzazioni, è necessario partire dalla comprensione dell'area che si vuole rappresentare e riorganizzare, ricorrendo a diverse fasi di analisi come l'identificazione e la qualificazione dell'ambiente del Sistema Informativo (SI), la definizione di tutte le parti interessate e la valutazione delle necessità e degli sviluppi futuri a breve e lungo termine.

In considerazione della portata della problematica di valorizzazione dei documenti e delle conoscenze della DIN, è stato obbligatorio effettuare un audit approfondito del SI della struttura che ci ha consentito anzitutto di avere una panoramica del volume della documentazione e della complessità del sistema. Ma l'intento è di andare oltre l'audit per affrontare temi e problematiche dal punto di vista della Direzione e degli utenti. Le osservazioni di questi ultimi tendono generalmente a considerare i malfunzionamenti del sistema come indici di problemi di livello superiore² che è necessario capire tramite un'analisi più approfondita, contestualizzando il sistema. Il metodo di analisi scelto ha permesso che il confronto delle osservazioni con i sondaggi presso gli utenti desse la certezza che documenti e utenti sono legati, all'interno dell'azienda, dalla nozione strutturante di *core business*. Tali risultati ci hanno orientato verso la valorizzazione del patrimonio informativo attraverso la competenza tecnica dell'organizzazione.

2.2. Il funzionamento del Sistema Informativo documentale attuale

I metodi di audit dei SI sono numerosi ma non corrispondono sempre al contesto industriale e non unicamente informatico³. Si è scelto di basare la diagnosi del sistema sul metodo di gestione

² M. DURAMPART, *L'outil donne à voir l'organisation*, in "Actes DOCSI-ERSICOM", Lyon, France 2006.

³ A. CARLIER, *Stratégie appliquée à l'audit des systèmes d'information*, coll. "Systèmes d'information", 2 éd., Hermès, Paris 1994; FMOI, *Approche méthodologique pour identifier les besoins en information des ingénieurs*, UNESCO, Paris 1994; E. SUTTER, *L'évaluation et les indicateurs de la performance*, ADBS, Paris 2006; F. ZANIER, *L'élaboration d'un tableau de bord : comment évaluer un centre de documentation*, ADBS, Paris 1995.

documentale SCRIP (Strategie, Criteri e Rappresentazione dell'Informazione Professionale) di Jean Pinteau⁴ e sulla norma francese AFNOR FD X50-185⁵. L'osservazione dei diversi livelli del sistema documentale – “sovra-documentale” (la funzione documentale nel suo contesto globale), “inter-documentale” (sistemi e parti interessate), “intra-documentale” (il documento) - serve a sottolineare la modalità operativa dominante all'interno di un'azienda.

In contesti simili a quello della DIN di EDF, la modalità operativa è generalmente detta sequenziale (o anche lineare o diacronica) ed è caratteristica dell'influenza diretta di una gestione della qualità obbligatoria e onnipresente, centrata sul controllo delle varie fasi del documento, ovvero l'identificazione degli utenti, la tracciabilità e la possibilità di ripristinare lo stato documentale precedente. La modalità sequenziale è presente nel contesto di aziende che hanno una gerarchia chiaramente discendente, frequenti monitoraggi con validazioni multiple e quindi iter molto lunghi. In generale, si può trovare in ambienti che favoriscono i processi di sicurezza a discapito della velocità di intervento. Le esigenze normative, in un'azienda come EDF con un contesto informatico rilevante, sono inoltre in linea con questa modalità.

Per ovviare alla scarsa rapidità di risposta imposta da questa situazione, si moltiplicano strumenti ausiliari più o meno ufficiali, come i database Note, Access, Intranet, ecc., che, su iniziativa degli stessi utenti ed in modo relativamente anarchico, sfuggono a questa logica, basandosi su un funzionamento detto simultaneo (o sincronico). L'orientamento puramente *core business* e *project management* di questi strumenti, fanno sì che essi si focalizzino su reattività, cooperazione, trattamento delle informazioni molto più che sulla produzione di documenti. Il funzionamento sincronico è di solito la modalità che dà maggiormente l'impressione di caos documentale e il numero di database di tipo Note o Intranet costruiti all'interno di multinazionali come EDF può diventare considerevole.

Per esempio, alla DIN lo strumento ufficiale è la GED (Gestione Elettronica Documentale), che contiene circa 10 milioni di documenti, al punto da diventare talmente pesante che possono essere necessari anche 15 minuti per ottenere il risultato di una semplice richiesta. Per di più, la GED non è stata pensata per una ricerca tramite parole chiave: si può cercare un documento solo con il suo riferimento esatto (che si deve quindi conoscere in anticipo). Questa modalità di ricerca non costituisce un problema per i senior dell'azienda, che sono anche gli autori di tutta la documentazione tecnica fin dalla costruzione delle centrali nucleari francesi negli anni Settanta, ma per i nuovi ingegneri che non conoscono la documentazione di riferimento, è un ostacolo enorme al lavoro quotidiano. Nonostante l'installazione di uno strumento di *flow research* (di tipo Google) per assistere la ricerca di informazioni nella GED, gli utenti sono costretti a scavalcare il sistema creando i loro propri strumenti paralleli. Nella DIN si contano una moltitudine di database e *a fortiori* nel Gruppo EDF. I formati sono molto diversi: Note, Oracle, Access, Excel, Filemaker, ecc. Prima della separazione tra EDF e GDF (Gaz de France) esistevano più di 20.000 database per le due entità, di cui 400 solo per la DIN, e più di 162 Intranet in EDF.

In realtà si tratta di una normale evoluzione del funzionamento lineare, sotto il peso dell'obbligo di consultazione, della necessità di rapidità di risposta e di lavoro in tempo reale da parte degli utenti che si applicano generalmente a materiale volatile e mutevole (documenti di gestione di progetto, e-mail, annotazioni, ecc.). La modalità sincronica si applica di solito a un contesto di competenze generalizzate il cui livello è piuttosto omogeneo. Nel contesto analizzato, il 70% degli utenti sono quadri/ingegneri che lavorano essenzialmente tramite una rete documentale informale e di e-mail, con un contenuto di informazioni che va oltre la GED ufficiale. Gli utenti rispondono al bisogno di consultazione e di risposte rapide richieste dalle attuali pratiche di lavoro mediante lo sviluppo di questi sistemi paralleli. Quindi emergono due SI: uno ufficiale e uno officioso, sviluppati di pari passo in direzioni e con ritmi diversi. Questa situazione si è verificata nel caso in esame ed è peggiorata progressivamente, creando dei veri e propri fallimenti nella gestione delle informazioni aziendali.

⁴ J. PINTEAU, *Reengineering des systèmes documentaires*, Les éditions d'organisation, Paris 1995.

⁵ AFNOR, *Norme FD X50-185 Management de l'information*, Afnor, Paris 2004.

Tali modalità operative e le loro caratteristiche, siano esse sequenziali o sincroniche, per poter essere equilibrate, devono essere correlate ai rispettivi problemi e obiettivi, ma anche ai mezzi e alle esigenze di documentazione degli utenti. Il conflitto tra questi elementi genera automaticamente malfunzionamenti che possono influire sul sistema documentale nel suo complesso e comportarne il rifiuto di utilizzo da parte degli utenti, come avviene nel nostro caso. Tutto questo conferma, in generale e nello specifico caso oggetto di studio, la necessità di valorizzare il patrimonio documentale dell'azienda e le conoscenze che contiene bilanciando le due modalità operative del Sistema Informativo.

2.3. La questione degli utenti e dei contesti trasversali

Allineare le risorse con i bisogni è stato il primo obiettivo. In termini di utenti, i problemi di accessibilità ai documenti sono concentrati nella GED: modalità di ricerca e interfaccia sono poco adatte all'esigenza finale. L'attenzione è stata focalizzata sul fatto che le informazioni di riferimento per i *core business*, così difficili da sfruttare, sono un patrimonio fondamentale per l'azienda. Tuttavia questo patrimonio è diventato nel tempo sparpagliato (database, GED, biblioteca, Internet, archivi personali, ecc.), eterogeneo (formato cartaceo o digitale, immagini, testi, e-mail, link Internet, ecc.) e difficilmente accessibile per tutti i documenti di gestione corrente dell'azienda (mescolati indistintamente nella GED con ogni genere di documento).

D'altra parte, alcuni contesti socioeconomici impongono oggi a molte aziende dei vincoli trasversali che richiedono di utilizzare la documentazione in modo altamente operativo. I problemi sono numerosi e critici. Il primo è il rinnovamento delle competenze: l'uscita in massa degli esperti che hanno vissuto la costruzione delle centrali nucleari francesi negli anni Settanta non sempre coincide con l'integrazione dei nuovi arrivati. Ciò comporta una significativa perdita di conoscenze tecniche nel momento di intersezione di due generazioni di tecnici. Questa perdita è particolarmente importante in una società dalla storica tradizione di *compagnonnage* (formazione professionale della durata di 1-2 anni). La questione del rinnovamento delle competenze rientra direttamente nella problematica documentale, ossia nella perpetuazione della conoscenza del *core business*. A questo si aggiunge il problema della vita degli impianti: le centrali nucleari hanno raggiunto i trent'anni di attività previsti alla loro costruzione. La documentazione di riferimento gioca quindi un ruolo vitale per sostenere la graduale sostituzione delle strutture e supportare la concezione dei reattori di nuova generazione (EPR: European Pressurized Reactor). Infine, è d'obbligo citare il tema, costante in ingegneria nucleare, della trasparenza presso l'Autorità di Sicurezza Nucleare europea. In conclusione, la documentazione di riferimento dei *core business* è il fulcro dell'attività di ingegneria nucleare e quindi oggi deve essere necessariamente classificata, perpetuata e valorizzata.

Pertanto, la contestualizzazione del SI ha permesso di trarre conclusioni a livello globale. In prima battuta, è necessario omogeneizzare le pratiche di informazione attraverso la standardizzazione della classificazione nella GED ufficiale e a livello delle relative pratiche d'uso in contrapposizione alla moltitudine di database. Poi, si dovrà semplificare il sistema spostandosi verso la ricerca di informazioni più che su quella di documenti. Infine, sarà necessario avvicinarsi alle pratiche degli utenti.

Detto questo, gli obiettivi del presente lavoro sono quelli di rendere la documentazione *core business* più fruibile e coerente e di esaminare sia gli aspetti gestionali che di ricerca delle informazioni attualmente mancanti. Lo scopo è quindi quello di valorizzare, perpetuare e facilitare la trasmissione delle conoscenze tecniche che la documentazione contiene, ovvero l'unica traccia costante lasciata nell'azienda dai vari ingegneri che ne sono stati parte (pensionamenti, trasferimenti, ecc.): gli artefatti umani creatori di conoscenze passano, gli artefatti documentari vettori di conoscenze rimangono.

3. Verso un'organizzazione coerente degli artefatti documentali vettori di conoscenze tecniche

3.1. *Inquadramento del dominio e inizio di riflessione verso una base di conoscenza core business*

Si è scelto di basare il presente lavoro su una delle competenze sensibili dell'azienda, gli “incidenti nucleari gravi”, in cui il bisogno di intervento era più importante che altrove. Il primo passo del metodo è stato quello di localizzare le competenze e di creare una rete di esperti del settore, ossia diciannove ingegneri, che sono stati scelti con vari profili e livelli di abilità. Ognuno ha, difatti, una specialità nel settore. Sono state condotte interviste aperte con loro per capire i flussi di conoscenze in questo ambito di competenza (come si formano, si divulgano, circolano e le fonti) e come sono articolati *core business* e documenti. Le conoscenze da individuare in questo caso sono quelle del percorso professionale che ha portato gli ingegneri al livello di competenze richiesto dal settore degli “incidenti gravi” e quelle apprese nel tempo all'interno di tale settore. Queste conoscenze pratiche si basano sull'esperienza e sono anche chiamate conoscenze contestuali. Sono quelle di interesse per i nostri scopi, in un sistema di *compagnonnage*, e sono complementari alle conoscenze concettuali acquisite studiando ingegneria.

A livello documentale, le interviste hanno confermato alcune importanti necessità: accedere alle informazioni per argomento o per navigazione tra i *core business* dell'azienda e nei testi; agevolare l'accesso dei nuovi arrivati alla competenza; e infine, soddisfare i diversi profili di utenti. A livello delle conoscenze, le interviste con gli ingegneri hanno anche evidenziato problemi di accesso alla competenza in assenza di senior che possono aiutare i nuovi assunti, ai quali viene direttamente richiesto di lavorare su una tematica precisa (per esempio: il rischio idrogeno). Poiché questo tipo di ricerca è impossibile nella GED, gli utenti si orientano verso la ricerca di informazioni esterne o informali, per esempio attraverso Google, il reperimento di studi di ex allievi, la rete esterna presso altre aziende, ecc.

Gli step seguenti sono stati guidati principalmente dalla volontà aziendale di invertire la logica che in precedenza sosteneva i progetti di SI all'interno dell'azienda partendo dalle pratiche reali degli ingegneri. Pertanto, si è ritenuto che il SI (che è la porta di accesso alle conoscenze del dominio) più gestibile per un utente sia quello costruito seguendo ovviamente i suoi usi, ma soprattutto la sua logica (quella che guida la costruzione delle conoscenze di perizia nel dominio) e il suo linguaggio (che sostiene la costruzione di tali conoscenze). L'intento era quindi di rendere facilmente accessibile la documentazione e le conoscenze che essa contiene, tramite la terminologia del settore.

Per Lamizet⁶, l'azienda è, infatti, uno spazio pubblico da considerare come uno spazio di comunicazione in cui circolano le conoscenze collettive che vi vengono costruite. Queste conoscenze sono mutate attraverso i documenti che ne diventano la rappresentazione sociale. Il linguaggio proprio degli ingegneri tiene vive, aggiorna e oggettiva le conoscenze collettive. Il linguaggio *core business* è quindi la chiave per promuovere le conoscenze tecniche.

Quindi, l'approccio scelto, sintetizzato nel diagramma in basso, parte dal postulato che le conoscenze in ingegneria nucleare passano attraverso un *medium* rappresentato dai documenti tecnici che sono l'espressione di concetti propri dei *core business*. Attraverso la mappatura delle conoscenze nate dalle abilità individuali e collettive, sarà possibile valorizzare la conoscenza e i *core business* che la concretizzano, per gestire meglio il patrimonio informativo dell'azienda. Questa mappatura è volta a realizzare una rappresentazione esplicita in una base di conoscenza, ponendo l'accento sul “*business sense*” dell'organizzazione.

⁶ B. LAMIZET, *Les lieux de la communication*, Mardaga, Liège 1992.

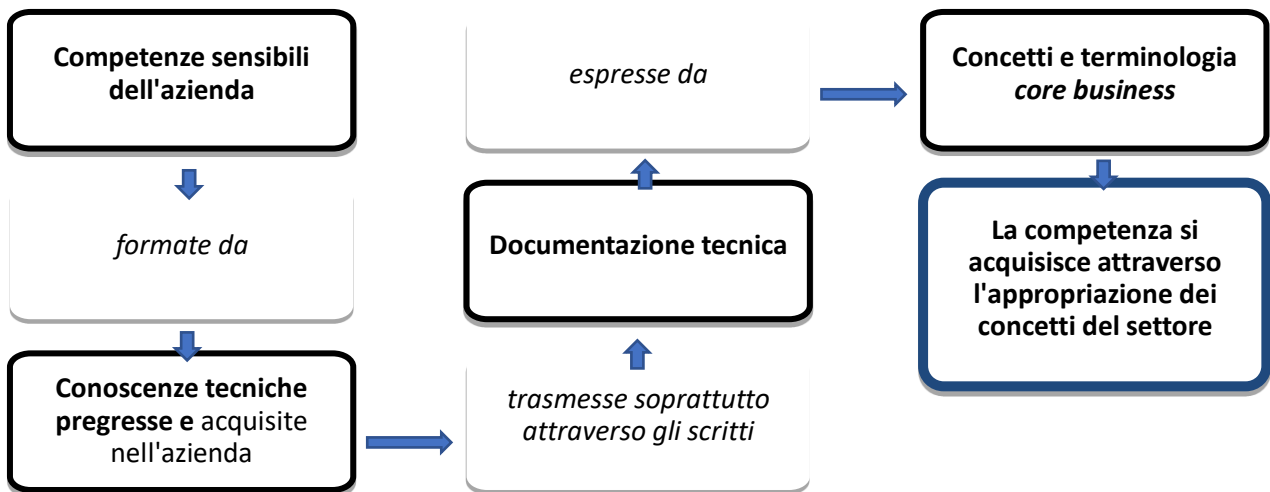


Figura 1. Mediazione della conoscenza.

3.2. Organizzazione delle conoscenze e della documentazione di riferimento nel dominio

Gli “incidenti gravi” sono un settore di competenze multidisciplinari al limite di tutti i fenomeni fisici. In generale, acquisire queste competenze richiede conoscenze generali, poiché non esiste un addestramento esterno specifico per gli “incidenti gravi”, in termoidraulica, meccanica, materiali, chimica e, in misura minore, neutronica. Le interviste con gli esperti hanno permesso di osservare in primo luogo, che si ritrovano costantemente gli stessi temi chiave richiesti per acquisire le competenze; in secondo luogo, che il lavoro (specializzazione e *project work*) e la formazione delle conoscenze degli ingegneri si organizzano seguendo il corso fenomenologico o fisico di un “incidente grave”. Questo corso fenomenologico ha luogo secondo sequenze logiche, proprio come la logica che sostiene la costruzione della conoscenza. La stessa organizzazione dei fenomeni fisici degli “incidenti gravi” ha indicato la procedura da seguire per costruire la mappatura delle conoscenze collettive di settore. Ciò ha permesso di prendere in considerazione la possibilità di accesso a questo tipo di competenze attraverso i riferimenti collettivi al dominio.

Poi, le interviste hanno rivelato che determinate fonti di informazione sono riconosciute utili dai *core business* per l'accesso alla competenza specifica di dominio al fine di acquisirne le conoscenze richieste. Si tratta di una serie di documenti interni, che è stato possibile individuare con precisione e che trattano in modo molto generale di: incidenti, operazioni, incidenti gravi, fenomenologia e sicurezza nucleare. I serbatoi di informazioni, ovvero dove reperire questi documenti di riferimento per l'accesso alle competenze, sono principalmente gli armadi (che rimangono negli uffici e non seguono i loro proprietari, in particolare in caso di pensionamento o trasferimento: contengono una serie di documenti di riferimento che non sono elencati da nessuna parte), il sito Web europeo Sarnet dedicato agli incidenti gravi in aree specifiche, un database interno e un'intranet. I documenti di riferimento identificati durante le interviste sono circa 700, sparsi ed eterogenei. Sono stati riuniti per formare una Raccolta specifica sugli *incidenti gravi*. Questi documenti non erano molto sfruttabili, essendo spesso di accesso limitato e quasi tutti in formato immagine. Sono stati quindi trattati via OCR (Optical Character Recognition), al fine di renderli accessibili per l'accesso alle competenze di dominio. Lo scopo era di offrire una guida operativa ai *core business*, strutturandone le conoscenze tramite la loro terminologia propria e questo, mediante la costruzione di una base di conoscenze che comprende un'ontologia e una terminologia di dominio, costruite con gli esperti, che sono al centro del presentemetodo. Questa base di conoscenze poggerà sulla Raccolta di documenti di riferimento del dominio, in formato testo, coerente e mirata.

4. Fondamenti termino-ontologici

4.1. Fissare la lingua di specialità tramite la terminologia

Il *Treccani* ci dà una definizione della terminologia (*comp. del lat. terminus e dal gr. logos, scienza*)
1. *L'insieme dei termini e delle espressioni propri di una scienza, di un'arte, di una tecnica o di una qualsiasi disciplina o attività specialistica.* 2. *Disciplina che ha come oggetto lo studio teorico della denominazione di concetti e realtà appartenenti alle diverse branche del sapere e in partic. della scienza e della tecnica.*

La terminologia, rappresentazione della “scienza” di un determinato dominio, è in effetti uno strumento che fissa il linguaggio specializzato di un campo sociale e garantisce riferimenti affidabili attraverso riferimenti stabili. Dirige «l'interpretazione che deve essere associata a ciascun termine e definisce il termine da utilizzare per designare un concetto o un oggetto»⁷. La consensualità attorno a questa terminologia, intrinsecamente legata a un contesto specifico, garantisce così una buona trasmissione, ricezione e appropriazione delle informazioni (o conoscenze) che trasmette nel suo campo sociale o contesto. Di fronte all'ambiguità del linguaggio quotidiano, i problemi di interpretazione sono quindi considerevolmente ridotti, poiché si limitano a un determinato dominio in cui l'uso è stabilizzato, ma persistono ancora su questioni relative all'evoluzione del linguaggio nel tempo o comunità d'uso, com'è il caso delle comunità *core business* che formano il dominio “incidenti nucleari gravi”. È possibile in effetti osservare differenze di significato da una comunità d'uso a un'altra, all'interno dello stesso dominio, per un sostantivo o un'espressione fissa.

Lo scopo di creare terminologie in campo industriale o scientifico è quello di eliminare questi problemi intrinseci al linguaggio umano, in un campo specifico la cui delimitazione rende più semplice la costruzione di un linguaggio standardizzato. Si tratta quindi di fissare gli usi, di formalizzarli più vantaggiosamente per alleviare il più possibile i problemi sopra menzionati. La vocazione di una terminologia è prescrittiva. La terminologia è quindi di per sé un linguaggio di specialità. I termini che formano questa terminologia sono specifici a un *core business* e si riferiscono a una «realtà extralinguistica condivisa: oggetti (artefatti), pratiche, metodi di processo, ecc. Questi termini non hanno lo stesso status delle parole della lingua comune»⁸. La nozione di normalizzazione è legata a quella della terminologia proprio dall'origine della parola, poiché “terminus” in latino indica quello che limita.

Questa necessità di standardizzazione dei termini scientifici e tecnici e del loro significato emerge nel diciottesimo secolo, era dell'enciclopedia e del pensiero che una scienza sia soprattutto un linguaggio ben fatto. Ad esempio, la logica di Port Royal già ipotizzava che il linguaggio perfetto si basa su un principio di concetti perfettamente definiti. Per Etienne Bonnot de Condillac, prete e filosofo le cui opere si adattano a questo movimento, le parole non potevano essere arbitrarie, ma dovevano riflettere esattamente le cose. Inoltre affermava: «Poiché le parole sono i segni delle nostre idee, il sistema linguistico deve essere formato su quello delle nostre conoscenze»⁹. Ma la terminologia come disciplina prende davvero forma nella prima metà del XX secolo, con Eugen Wüster, il quale ne fissa i principi metodologici. Nelle sue opere fondatrici¹⁰, si riferisce alla terminologia come «la ricerca di un sistema di forma neutra, un simbolismo purificato delle scorie delle lingue storiche»¹¹. Si tratta quindi di costruire un linguaggio artificiale, il meno ambiguo possibile, di strutturare un campo di conoscenza. Verso la fine degli anni Ottanta, i linguisti hanno preso la posizione della natura puramente linguistica del termine e la costante connessione del suo studio ai testi. Su questo punto, i progressi dell'IA (Intelligenza Artificiale) hanno, come vedremo nel

⁷ S. LAINE-CRUZEL, *Valoriser l'information : enrichir les systèmes et les interfaces, faciliter l'appropriation*, in “Actes Docsi-Ersicom”, Lyon, France, 2006b.

⁸ C. ROCHE, *Terminologie et ontologie*, in “Langages”, n. 157, 2005, pp. 48-62.

⁹ E. BONNOT DE CONDILLAC, *Cours d'étude pour l'instruction du Prince de Parme, Tome Premier, Grammaire*, Du Villard Fils & Nouffer, Genève 1780.

¹⁰ E. WÜSTER, *L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses*, in *Textes choisis de terminologie*, edited by G. Rondeau et H. Felber, Girsterm, Université de Laval, Québec 1981, pp. 55-108.

¹¹ A. SOULEZ, *Manifeste du Cercle de Vienne et autres écrits*, PUF, Paris 1985.

presente lavoro, sconvolto la disciplina e consentono ora trattamenti interessanti su grandi volumi di testi.

4.2. Gli usi nella lingua

Nel modello triadico del significato linguistico si possono considerare tre entità che legano il nostro campo di conoscenze alle sue espressioni linguistiche. Tale modello trova le sue basi nel *Peri hermeneias* di Aristotele¹², che alcuni filosofi medievali come Tommaso d'Aquino, hanno interpretato sotto la forma *Vox - Conceptus - Res*. Ma la riflessione contemporanea ispirata da Peirce¹³, lo rappresenta come tale da Ogden e Richards¹⁴ *Symbol – Referent – Thought or Reference* (dove il *Symbol* può essere paragonato al *Significante* di Ferdinand de Saussure e il *Thought or Reference* al *Significato* di Saussure¹⁵).

Sulla base di questo triangolo semiotico, Roche¹⁶, che ha collaborato a questa parte termino-ontologica del presente lavoro, presenta la teoria su cui si basa il nostro approccio, di tre linguaggi che si riferiscono a tre pratiche.

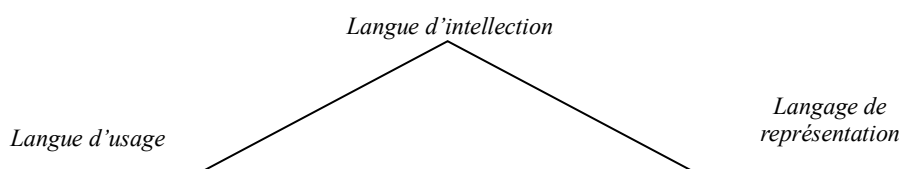


Figura 2. I tre linguaggi e le tre pratiche corrispondenti secondo Christophe ROCHE.

La *langue d'usage* in linguistica è legata ad un uso pratico dell'universo e tiene conto della nozione di locutore.

La *langue d'intellection* è legata alla conoscenza intellettuale che noi, osservatori, abbiamo dell'universo rispetto a principi epistemologici. Essa è indipendente delle contingenze legate agli usi. I *langages de représentation* sono principalmente degli strumenti che hanno la finalità di rappresentare le entità del mondo reale per manipolarle.

Queste tre entità che formano il sistema complesso di espressione delle nostre conoscenze devono essere viste come i tre livelli che ci conducono dalla terminologia al suo sfruttamento computazionale.

I discorsi scientifici o tecnici appartengono alla lingua di specialità. Sono la parte più visibile di un settore. In questo contesto abbiamo a che fare con termini d'uso che costituiscono la base di interpretazione e creazione di senso nell'atto del discorso. È quindi nello scambio intersoggettivo tra locutori agenti cognitivi, che assumono il loro senso, sia a livello dell'intenzione (produzione) che dell'interpretazione dei discorsi. Sono qui di interesse i rapporti tra *significanti* (i termini d'uso, la forma linguistica) e *significati* in un contesto particolare (la forma extralinguistica). Oltre l'identificazione pura dei termini, l'analisi del discorso può portare una certa conoscenza del sistema nozionale. Considerando che la terminologia è una lingua specializzata caratterizzata dall'uso di parole specifiche ad una pratica e che tali parole sono i soli dati osservabili che ci rimangono, sembra naturale basare la costruzione della terminologia sul trattamento degli artefatti dove sono fissati, quindi, i testi scientifici. Oggi numerosi lavori dimostrano l'interesse dell'estrazione di conoscenza o di terminologie sulla base di testi¹⁷. I *corpora* digitali sempre più accessibili e l'uso dell'informatica portano a risultati ricchi come nella semantica distribuzionale.

¹² ARISTOTELE, *De l'interprétation*, trad. di J. Tricot, Editions Les échos du maquis, Paris 2014.

¹³ C. S. PEIRCE, *Collected Papers*, Harvard University Press, Cambridge 1960.

¹⁴ OGDEN C.K., RICHARDS, L. A., *The Meaning of Meaning*, Routledge and Kegan Paul, London 1923.

¹⁵ F. (DE) SAUSSURE, *Cours de linguistique générale*, Payot, Paris 1916.

¹⁶ C. ROCHE, *Terminologie et ontologie*, cit.

¹⁷ G. A. MILLER et al., *Introduction to WordNet : an on-line lexical database*, in "International Journal of Lexicography 3", n. 4, 1990, pp. 235-244; B. HABERT, A. NAZARENKO, A. SALEM, *Les linguistiques de corpus*, Editions Armand Colin, Paris 1997; N. AUSSENAC-GILLES, A. CONDAMINES, *Entre textes et ontologies formelles : les bases de connaissances terminologiques*, in *Ingénierie des*

4.3. Scelta del metodo di costruzione di una base di conoscenza

Una *knowledge base*, o base di conoscenza, combina la conoscenza o la perizia di un'area specialistica, in una forma che può essere sfruttata da un computer. Nel caso in oggetto la base sarà costruita su un'ontologia e una terminologia del dominio per dare accesso alla documentazione di riferimento.

Le basi di conoscenza sono al centro di molte applicazioni, come la gestione del *know-how* e delle competenze, il monitoraggio, la ricerca di informazioni, ecc. La loro costruzione è un'operazione decisamente umana, intellettuale e complessa, per la quale non esiste una metodologia universalmente accettata¹⁸. Gli approcci sono spesso empirici, come nel caso in oggetto. Per descrivere correttamente il campo di studio preso in esame, l'intervento precedente i trattamenti informatici dei *corpora* tecnici è stato molto pesante, in quanto influisce fortemente sulla rilevanza dei risultati. Ma l'intervento umano è stato notevole anche dopo i trattamenti, per l'interpretazione, la convalida e il perfezionamento dei risultati proposti dai sistemi. Pertanto, tenuto conto di un SI complesso, è apparso molto più ragionevole adottare un approccio qualitativo che unisse in maniera complementare il lavoro con gli esperti e l'estrazione sulla base dei testi.

Tra i lavori del campo orientati verso la linguistica di corpus, guidati in Francia dal gruppo TIA (Terminologia e Intelligenza Artificiale), si trovano approcci principalmente semasiologici. Si citano per esempio le ricerche di Aussenac-Gilles e Condamines che propongono un metodo per la costruzione di ontologie basate su analisi di corpus, con tecniche di elaborazione automatica delle lingue, per produrre una base di conoscenza terminologica (BCT)¹⁹. Si trovano anche le ricerche condotte internamente al Gruppo EDF da Henri Boccon-Gibod²⁰ per la coerenza del SI documentale. Il metodo che è stato implementato segue una logica diversa. Considerando i lavori attuali, si è preferito avvicinarsi maggiormente alle ontologie regionali di Bernard Bachimont²¹ laddove, ad esempio, si prendono i testi come fonte di conoscenza e il modello è costruito sui testi e sulle esigenze della domanda di destinazione, senza un obiettivo di generalizzazione. Peraltro, i lavori di capitalizzazione delle conoscenze, condotti in EDF da Christophe Roche²², rappresentano ai fini del presente lavoro un approccio onomasiologico particolarmente interessante, anche se l'aspetto di gestione dei documenti è secondario. Il presente metodo si inserisce tra i precedenti, utilizzando trattamenti automatici, ma promuovendo l'intervento umano e ponendo essenzialmente l'esperto al centro del lavoro, in modo da bilanciare costantemente i due aspetti e ridurre le distorsioni che possono emergere tra i risultati di trattamenti automatici o basati esclusivamente sull'esperto.

5. La terminologia *business* per elaborare l'informazione tecnica

5.1. Lo strato linguistico: la *langue d'usage*

La base di conoscenza costruita è stata quindi concepita per consentire la valorizzazione della documentazione tramite il linguaggio, vale a dire la terminologia e i concetti specifici dei *business core*, per una migliore appropriazione delle informazioni e delle conoscenze. Il presente lavoro, basato interamente sullo scambio costante con gli esperti, si è svolto in diverse fasi ed è basato sulla

connaissances. Evolutions récentes et nouveaux défis, edited by J. Charlet et al., Eyrolles, Paris 2000; M. C. L'HOMME, *Évaluation de logiciels d'extraction de terminologie : examen de quelques critères*, Université de Montréal, Montréal 2000; BOURIGAUULT D., JACQUEMIN C., L'HOMME M. C., *Recent advances in computational terminology*, John Benjamins Publishing, Amsterdam/Philadelphia 2001; R. TEULIER, J. CHARLET, P. TCHOUNIKINE, *Ingénierie des connaissances*, L'Harmattan, Paris 2005; A. LAROCHE, P. DROUIN, G. BERNIER-COLBORNE, *Étude de l'influence de la taille du corpus de référence sur l'extraction terminologique automatique contrastive*, Université de Montréal, Montréal 2011.

¹⁸ LAINE-CRUZEL S., *Terminologie et intelligence artificielle*, in *Encyclopédie de l'informatique et des Systèmes d'Information*, Vuibert, Paris 2006a.

¹⁹ N. AUSSENAC-GILLES, A. CONDAMINES, *op.cit.*

²⁰ H. BOCCON GIBOD, *Application de méthodes et outils de Web sémantique pour la gouvernance d'un système d'information industriel*, EDF R&D, Clamart 2006.

²¹ B. BACHIMONT, *Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances*, in *Ingénierie des connaissances. Evolutions récentes et nouveaux défis*, edited by J. Charlet et al., Eyrolles, Paris 2000.

²² C. ROCHE, *Dire n'est pas concevoir*, in "Actes des 18es Journées Francophones d'Ingénierie des Connaissances", Grenoble, France 2007.

metodologia di Christophe Roche²³. In primo luogo, la definizione di un corpus di testi con l'esperto del settore, corpus che si è voluto molto concentrato, composto da otto testi tecnici (570 pagine) che coprono il 90% dei concetti del settore. Si è così voluto privilegiare la qualità sulla quantità di testi considerati rappresentativi.

Il passo successivo è stato l'estrazione automatica di termini dal corpus sotto forma di due diversi lessici, uno dei sintagmi nominali e l'altro delle sigle che sono ampiamente utilizzate nel linguaggio dell'ingegneria nucleare. Dopo un'importante elaborazione manuale, sono risultati 687 termini convalidati dagli esperti. La fase successiva del lavoro è consistita nella costruzione di una rete concettuale. La costruzione di questa rete direttamente dai lessici estratti, come nei metodi semasiologici, si è rivelata inestricabile e non pertinente per gli esperti. In effetti, nonostante le prestazioni del Trattamento Automatico della Lingua (TAL), l'incompletezza dei testi è una problematica sempre presente. La comprensione dei discorsi (mediati o no tramite i testi) implica quindi che i locutori condividano un universo extralinguistico comune. Ma come captare l'intenzione del destinatario e l'interpretazione che ne farà il destinatario quando i testi non le esprimono? Il significato preso in considerazione in tanti lavori come base nozionale riguarda il senso, che potrà variare nello stesso discorso e profondamente secondo gli usi. Questo punto di vista fornisce risultati non riutilizzabili, completamente dipendenti dal corpus, che non sono né sistemi nozionali (perché ancora nel sistema linguistico) né terminologie.

5.2. Lo strato extralinguistico: la langue d'intellection

Si vuole a questo punto uscire dalla lingua, distanziandosi dai lessici e seguendo un approccio onomasiologico (basato sulla consultazione periodica degli esperti), per accedere al significato dei termini, indipendentemente dagli usi. Esso verrà definito allora come un senso normato, comune ai diversi discorsi, con riferimento ad una realtà extralinguistica. Roche definisce «il senso di una parola come il significato attualizzato nel discorso»²⁴. Questo strato superiore non è reperibile nei testi perché non ci sono assolutamente concetti, ma solo tracce linguistiche dei loro usi. Si entra qui nella nozione di lingua d'intellection. La definizione di questa realtà potrà, a sua volta, farsi tramite un linguaggio di rappresentazione, per esempio concettuale, che sarà sviluppato più avanti.

Secondo Helmut Felber «è opportuno ricordare che ogni lavoro terminologico dovrebbe basarsi sulle nozioni e non sui termini»²⁵: la terminologia si interessa alle parole solamente perché rappresentano delle nozioni (o concetti). Un termine è quindi la somma di un nome nella lingua e del concetto che gli corrisponde, la terminologia è quella di un sistema di termini e di una modellizzazione concettuale (o nozionale) che li riflette. Essa, oltre la lingua d'uso, è una lingua d'intellection che riguarda le relazioni concetti/oggetti, l'articolazione extralinguistico/linguistico. Questa visione della terminologia la sottrae al confine restrittivo di una normalizzazione concentrata sull'universale per prendere in considerazione la varietà degli usi all'interno di una comunità, nonché i particolarismi linguistici e concettuali. Integrare questa diversità non è un ostacolo per la realizzazione di una terminologia che risponda ai criteri imposti dal mondo industriale e dalle sue necessità. È necessario che sia al massimo consensuale nella definizione dei termini, coerente e evolutiva. Questo lavoro preliminare di astrazione dallo strato linguistico dei lessici verso lo strato extralinguistico, è stato interamente umano oltre che indispensabile. La rete concettuale finale è composta da diverse vedute complementari.

²³ C. ROCHE, *Terminologie et ontologie*, cit.

²⁴ C. ROCHE, *Terminologie et ontologie*, cit.

²⁵ H. FELBER, *Terminology Manual*, Infoterm, Vienne 1984.

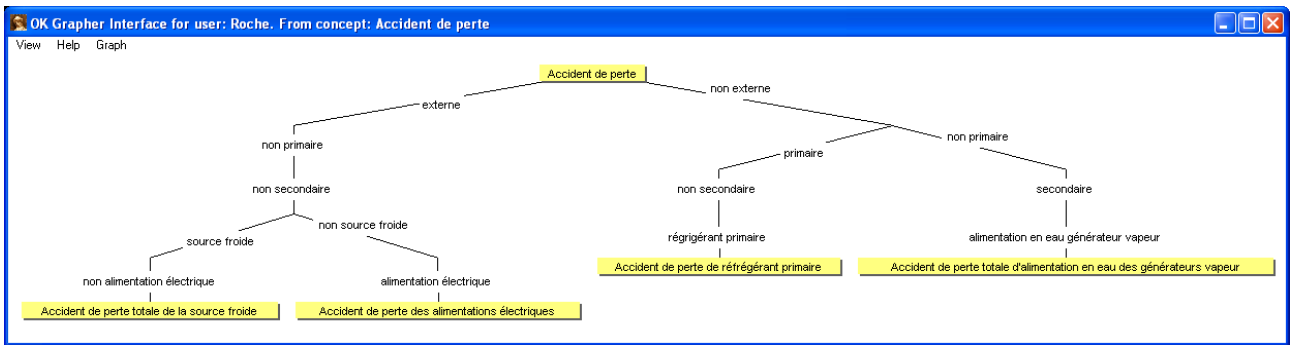


Figura 4. Dettaglio di *Événements initiateur*: ontologia *Accidents de perte*.

È stato a questo punto possibile unire la parte semasiologica e quella onomasiologica del lavoro. La terminologia è stata costruita collegando i termini dei due lessici con l'ontologia, per indicizzare i circa 700 documenti individuati attraverso le interviste con gli ingegneri. Questi documenti sono stati dematerializzati in formato testo per ottenere un formato uniforme. Il lavoro ha portato alla creazione di una piattaforma di test per la ricerca e la gestione delle informazioni tecniche, subordinata alla costituzione di un metodo che l'azienda può riutilizzare per mantenere o espandere l'approccio ad altre aree. Uno studio di feedback da parte degli utenti sarebbe stato particolarmente interessante, ma non ha potuto essere completato nel tempo previsto dal presente lavoro.

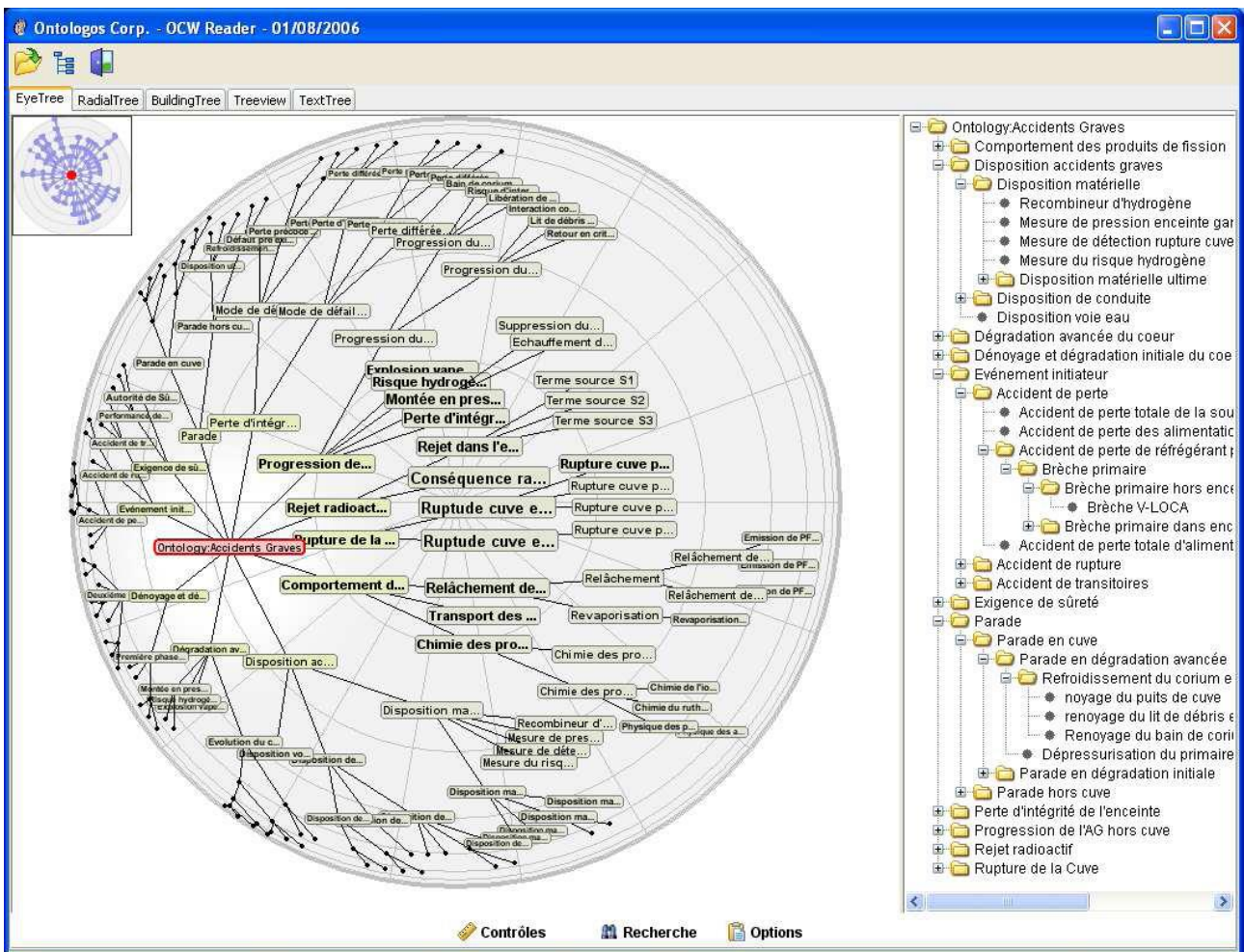


Figura 5. Piattaforma di test *Accidents Graves*.

6. Conclusioni

L'esempio offerto dal presente lavoro è solo uno tra i tanti. Oggi numerose aziende devono confrontarsi con queste problematiche. L'integrazione di massa dei computer negli ultimi quarant'anni, la valanga di informazioni, e oggi di dati, che ha indotto, hanno certamente facilitato alcuni aspetti delle professionalità tecniche, ma le aziende non sono riuscite in genere a mantenere il passo. La fiducia cieca nello strumento tecnico ed il rifiuto reiterato di investire in risorse umane adeguate a rendere questi strumenti intelligenti o almeno logici e controllabili, hanno portato alle situazioni complicate che è possibile osservare. In risposta, la tendenza è sempre più marcata: la forza lavoro qualificata in campo documentale continua ad essere ristretta e le risposte vengono sempre chieste agli strumenti.

Investire in fase iniziale nello strumento è di fatto essenziale. Consente l'analisi, tenendo conto di tutti i parametri coinvolti nei vari progetti. Ma questa fase è spesso evitata per ragioni di risparmio di tempo e denaro. Tuttavia, paradossalmente, scavalcare questa costruzione preliminare produce in realtà a posteriori, notevoli costi aggiuntivi con risultati finali inadeguati. Il fattore umano non viene preso in considerazione: gli utenti sono ignorati sia nelle loro richieste iniziali che per quanto riguarda l'uso finale che faranno degli strumenti. Viene chiesto loro di adattarsi allo strumento e non viceversa. L'informazione è privata di qualsiasi valore aggiunto nonché della sua funzione di servizio.

Tuttavia, i problemi relativi alla valorizzazione delle informazioni e delle conoscenze tecniche dei *core business* andranno a crescere in molte strutture simili. Come abbiamo visto, i contesti tecnici, demografici, competitivi, ecc., porteranno ad un aumento sensibile di queste problematiche. Ma non è più possibile negare e rimandare il problema, lasciando accumulare sempre più le informazioni. In ogni caso, non potrà essere realizzato niente di costruttivo fintanto che lo strumento e non l'umano sarà al centro della riflessione. Anche il presente approccio usa lo strumento, ma propone un cambiamento di prospettiva fondamentale che prevede di concentrarsi prioritariamente sull'appropriazione delle informazioni e delle conoscenze da parte degli utenti. Pertanto gli esperti hanno un ruolo centrale nel metodo presentato. Obiettivo del presente lavoro è stato valorizzare il patrimonio documentale delle organizzazioni e le conoscenze tecniche che contiene, e renderli accessibili e utilizzabili dagli utenti utilizzando la propria terminologia. È stato realizzato mediante la costruzione di un'ontologia applicata ad un settore specifico. Basandosi su fondamenti termino-ontologici, il presente lavoro ha proposto un equilibrio tra i soli trattamenti automatici degli approcci semasiologici della linguistica di corpus e gli approcci onomasiologici basati esclusivamente sugli esperti. L'ontologia e la terminologia di dominio risultanti sono state utilizzate per gestire la documentazione tecnica di riferimento, per capitalizzare la competenza degli esperti e per trasferirne le conoscenze alle nuove generazioni di ingegneri.

Riferimenti bibliografici

AFNOR, *Norme FD X50-185 Management de l'information*, Afnor, Paris 2004.

ARISTOTE, *De l'interprétation*, trad. Di J. Tricot, Editions Les échos du maquis, Paris 2014.

AUSSENAC-GILLES N., CONDAMINES A., *Entre textes et ontologies formelles : les bases de connaissances terminologiques*, in *Ingénierie des connaissances. Evolutions récentes et nouveaux défis*, edited by J. Charlet et al., Eyrolles, Paris 2000.

BACHIMONT B., *Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances*, in *Ingénierie des connaissances. Evolutions récentes et nouveaux défis*, edited by J. Charlet et al., Eyrolles, Paris 2000.

BOCCON GIBOD H., *Application de méthodes et outils de Web sémantique pour la gouvernance d'un système d'information industriel*, EDF R&D, Clamart 2006.

BONNOT DE CONDILLAC E., *Cours d'étude pour l'instruction du Prince de Parme, Tome Premier; Grammaire*, Du Villard Fils & Nouffer, Genève 1780.

BOURIGAULT D., JACQUEMIN C., L'HOMME M. C., *Recent advances in computational terminology*, John Benjamins

Publishing, Amsterdam/Philadelphia 2001.

CARLIER A., *Stratégie appliquée à l'audit des systèmes d'information*, coll. "Systèmes d'information", 2^o éd., Hermès, Paris 1994.

DURAMPART M., *L'outil donne à voir l'organisation*, in "Actes DOCSI-ERSICOM", Lyon, France, 2006.

FELBER H., *Terminology Manual*, Infoterm, Vienne 1984.

FMOI, *Approche méthodologique pour identifier les besoins en information des ingénieurs*, UNESCO, Paris 1994.

GANDON F., DIENG-KUNTZ R., *Ontologie pour un système multi-agent dédié à une mémoire d'entreprise*, in *Ingénierie des connaissances*, L'Harmattan, Paris 2005.

HABERT B., NAZARENKO A., SALEM A., *Les linguistiques de corpus*, Editions Armand Colin, Paris 1997.

LAINÉ-CRUZEL S., *Terminologie et intelligence artificielle*, in *Encyclopédie de l'informatique et des Systèmes d'Information*, Vuibert, Paris 2006a.

LAINÉ-CRUZEL S., *Valoriser l'information : enrichir les systèmes et les interfaces, faciliter l'appropriation*, in "Actes Docsi-Ersicom", Lyon, France, 2006b.

LAMIZET B., *Les lieux de la communication*, Mardaga, Liège 1992.

LAROCHE A., DROUIN P., BERNIER-COLBORNE G., *Étude de l'influence de la taille du corpus de référence sur l'extraction terminologique automatique contrastive*, Université de Montréal, Montréal 2011.

L'HOMME M. C., *Évaluation de logiciels d'extraction de terminologie : examen de quelques critères*, Université de Montréal, Montréal 2000.

MILLER G. A. et al., *Introduction to WordNet: an on-line lexical database*, in "International Journal of Lexicography 3", n. 4, 1990, pp. 235-244.

OGDEN C.K., RICHARDS, L. A., *The Meaning of Meaning*, Routledge and Kegan Paul, London 1923.

PEIRCE C. S., *Collected Papers*, Harvard University Press, Cambridge 1960.

PINTEA J., *Reengineering des systèmes documentaires*, Les éditions d'organisation, Paris 1995.

ROCHE C., *Terminologie et ontologie*, in "Langages", n. 157, 2005, pp. 48-62.

ROCHE C., *Dire n'est pas concevoir*, in "Actes des 18es Journées Francophones d'Ingénierie des Connaissances", Grenoble, France, 2007.

SAUSSURE (DE) F., *Cours de linguistique générale*, Payot, Paris 1916.

SOULEZ A., *Manifeste du Cercle de Vienne et autres écrits*, PUF, Paris 1985.

SUTTER E., *L'évaluation et les indicateurs de la performance*, ADBS, Paris 2006.

TEULIER R., CHARLET J., TCHOUNIKINE P., *Ingénierie des connaissances*, L'Harmattan, Paris 2005.

ZANIER F., *L'élaboration d'un tableau de bord : comment évaluer un centre de documentation*, ADBS, Paris 1995.

WÜSTER E., *L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses*, in *Textes choisis de terminologie*, edited by G. Rondeau et H. Felber, Girsterm, Université de Laval, Québec 1981, pp. 55-108.