



UNIVERSITÀ DI PISA

**DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA DEI SISTEMI
DEL TERRITORIO E DELLE COSTRUZIONI**

**RELAZIONE PER IL CONSEGUIMENTO DELLA
LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE**

Mapping and Classifying Knowledge Management Tools with Text Mining.

SINTESI

RELATORI

IL CANDIDATO

Prof. Ing. Gualtiero Fantoni, Ph.D
Dipartimento di Ingegneria Civile e Industriale

Fabio Malluzzo
fabiomalluzzo@gmail.com

Dr. Vito Giordano
Dipartimento di Ingegneria dell'Informazione

Dr. Filippo Chiarello, Ph.D
*Dipartimento di Ingegneria dell'Energia
dei Sistemi del Territorio e delle Costruzioni*

Sessione di Laurea Magistrale del 28/04/2021

Sommario

Il *Knowledge Management (KM)* è l'insieme delle metodologie, processi e tecnologie, atti alla gestione e condivisione della conoscenza all'interno di un'organizzazione. In particolare, i metodi, hanno una importanza centrale nelle aziende, considerata la grande mole di informazioni che oggi devono gestire. In Letteratura non è presente né una precisa definizione e né classificazione di tali metodi, limitando la loro usabilità e comprensione.

Il lavoro di tesi mira a risolvere tale gap, offrendo un approccio quantitativo all'identificazione e alla classificazione dei metodi usati nel *KM*. L'identificazione è avvenuta sfruttando strumento di analisi automatica del testo usando come fonte Wikipedia. La classificazione poi si è basata sul definire le relazioni esistenti tra i metodi e sull'identificare il ruolo che hanno nel processo di *KM*.

Dall'analisi svolta sono stati identificati e classificati 245 strumenti, raggruppati in otto classi. Il risultato ottenuto può essere utile per orientarsi negli strumenti di *Knowledge Management* per i domini presi in considerazione.

Abstract

Knowledge Management (KM) is the set of methodologies, processes and technologies, acts to the management and sharing of knowledge within an organization. In particular, methods, they have a central importance in companies, considered the great amount of information that today must manage. In the literature there is no precise definition and classification of these methods, limiting their usability and understanding.

The thesis work aims to solve this gap, offering a quantitative approach to identification and the classification of the methods used in the *KM*. The identification took place using automatic text analysis tool using as a Wikipedia source. The classification was based on defining existing relationships between methods and identifying the role they have in the process of *KM*.

From the analysis carried out 245 tools were identified and classified, grouped in eight classes. The result obtained can be useful to orientate in the Knowledge Management tools for the domains taken into consideration.

1 Introduzione

Esiste allo stato dell'arte una generale mancanza di definizione e classificazione dei metodi utilizzati per svolgere le attività del processo di *Knowledge Management*. Ne consegue sia una limitazione nell'adozione dei metodi di gestione della conoscenza, sia una distorsione nel loro utilizzo.

L'obiettivo del mio lavoro di tesi è quindi di definire cosa sia un metodo per gestire la conoscenza e successivamente distinguere gli strumenti in base all'utilizzo e al ruolo che svolgono all'interno del processo.

Lo scopo è fornire uno studio utile a diversi stakeholder, ma in primis agli ingegneri gestionali, per riuscire ad individuare, identificare e utilizzare i metodi di gestione della conoscenza, in riferimento alle proprie esigenze, date dal raggiungimento degli obiettivi prefissati per ottenere un vantaggio competitivo.

Il primo passo dell'analisi consiste in una ricerca in Letteratura per riuscire definire il concetto di *Knowledge Management* e cosa siano gli strumenti/metodi di *Knowledge Management*.

L'analisi si suddivide in due macro-sezioni, una riguardante l'identificazione e l'altra la classificazione degli strumenti di *Knowledge Management*. L'identificazione consiste in una ricerca in Letteratura e successivamente, tramite Wikipedia, una espansione della ricerca. La classificazione consiste nell'assegnare gli strumenti identificati alle fasi del *Knowledge Management* e nel determinare le relazioni esistenti tra gli strumenti, utilizzando *Text Mining* per avere una mappatura automatica.

2 Revisione della letteratura

2.1 Processo di *Knowledge Management*

Il concetto principale del *Knowledge Management* è il termine conoscenza. La conoscenza è definita come informazione posseduta nella mente delle persone o acquisita da esperienze passate¹.

Il *Knowledge Management* ha a che fare con tre fattori: dati, informazioni e conoscenza. In sequenza di valore per una azienda, i dati sono fatti grezzi e non elaborati, le informazioni

¹ Si Xue, C. T. (2017). A Literature Review on Knowledge Management in Organizations. *Research in Business and Management*, 4(1), 30. <https://doi.org/10.5296/rbm.v4i1.10786>.

sono un insieme di dati organizzati ed elaborati e la conoscenza è un'informazione elaborata². Anche se la maggiore parte degli studi presenti in letteratura, fatica a fornire una definizione coerente ed utilizzabile di questi tre elementi, è chiaro che le relazioni tra dati, informazioni e conoscenza, dipende da come essi sono organizzati e interpretati. In particolare, i dati e le informazioni sono riferiti alla loro organizzazione e definizione, ad esempio definizione presente su manuali, mentre la conoscenza è ottenuta dall'interpretazione dei dati e dalle informazioni.

Una delle classificazioni del processo di *Knowledge Management* più utilizzate in Letteratura è la distinzione in quattro macro-fasi:

- **Creazione della conoscenza**, con lo scopo di generare conoscenza, e comprende le attività di creazione di idee, dati ed informazioni.
- **Identificazione della conoscenza**, con lo scopo di categorizzare le informazioni generate.
- **Organizzazione della conoscenza**, con cui le informazioni e i dati, vengono adattati alle esigenze interne dell'organizzazione.
- **Applicazione della conoscenza**, utilizzata per ottenere un vantaggio competitivo.

Questa struttura è risultata essere particolarmente utile per classificare gli strumenti di *Knowledge Management*.

2.2 Strumenti di *Knowledge Management*

Sono stati considerati strumenti di *Knowledge Management*, quelle metodologie e tecniche, utilizzate per svolgere le attività nei processi aziendali, tramite la manipolazione di variabili secondo la trilogia: prevedere, decidere e controllare³. Sulla base delle definizioni trovate in letteratura, si definiscono gli strumenti di gestione della conoscenza come tecniche per gestire la conoscenza dei processi aziendali o delle diverse funzioni di *business*, al fine di determinare un vantaggio competitivo. Ad esempio, nella fase di identificazione della conoscenza alcuni strumenti sono *PEST analysis*, *PERT*, *Quality function deployment*, in quanto sono definiti strumenti strategici che svolgono un ruolo di indentificare e mappare le informazioni e i dati nei processi aziendali.

² Si Xue, C. T. (2017). A Literature Review on Knowledge Management in Organizations. *Research in Business and Management*, 4(1), 30. <https://doi.org/10.5296/rbm.v4i1.10786>.

³ Martineau, R. (2017). What are management tools made of? The "listic" structure of managerial artifacts. *Management (France)*, 20(3), 239–265.

Una delle prime classificazioni presenti in Letteratura riguarda l'utilizzo dei sette strumenti di *Ishikawa*, utilizzati per far fronte ai problemi di affidabilità e controllo della qualità. Questi strumenti sono definiti come strumenti statistici che utilizzano modi di rappresentazione grafica facilitando la comprensione dei fenomeni presi in esame.

3 Metodologia

In questo capitolo vengono descritti i metodi per identificare e classificare gli strumenti di *Knowledge Management*.

3.1 Identificazione degli strumenti di *Knowledge Management*

Una ricerca in Letteratura degli strumenti utilizzati nel processo del *Knowledge Management* è il primo passo per identificare quali sono i metodi. L'identificazione inizia con una ricerca degli strumenti utilizzati nelle diverse *business functions*, come mostrato in Tab. 3.1⁴:

Tabella 3.1: Strumenti utilizzati per le diverse *Business Functions*

<ul style="list-style-type: none"> • Project management (<i>Work breakdown structure, PERT</i>) 	<ul style="list-style-type: none"> • Risk management (<i>Failure mode, effects, and criticality analysis; Fault tree analysis</i>)
<ul style="list-style-type: none"> • Business process management (<i>Business Model Canvas, Ansoff Matrix</i>) 	<ul style="list-style-type: none"> • Product management (<i>Material requirements planning, Master production schedule</i>)
<ul style="list-style-type: none"> • Strategic management (<i>SWOT analysis, PEST analysis</i>) 	<ul style="list-style-type: none"> • Quality management (<i>5M model, Checklist, Pareto analysis</i>)

Le entità trovate dalle fonti di ricerca, sono state confrontate con la definizione di strumento di *Knowledge Management* e la definizione presente su Wikipedia di ogni entità.

È stata utilizzata l'enciclopedia online Wikipedia, considerando le sue proprietà come fonte di conoscenza per l'estrazione automatica del testo, la delimitazione del campo e il clustering. Riuscendo sia ad individuare i metodi di gestione della conoscenza, che per identificare i legami tra tali metodi. Dalla ricerca iniziale, sono stati identificati 96 strumenti.

Successivamente è iniziata la fase di espansione, mediante l'estrazione degli *hyperlink* presenti nelle pagine Wikipedia degli strumenti, ottenendo un dizionario di strumenti di *Knowledge Management*⁵.

⁴ What Are the Branches of Business Management? - business.com. (n.d.). Retrieved February 10, 2021, from <https://www.business.com/articles/8-branches-of-business->.

Gli *hyperlink* sono collegamenti ipertestuali, nel nostro caso nelle pagine Wikipedia, che hanno lo scopo di condurre ad ulteriori unità informative a partire da una qualunque unità ad esse correlata⁶.

L'estrazione degli *hyperlink*, effettuata con tecniche di *Text Mining* implementate nel software R⁷, ha come primo scopo quello di riuscire ad espandere la lista degli strumenti, come secondo scopo quello di definire le relazioni esistenti tra gli strumenti. Si è iterato il processo di estrazione fino a raggiungere la saturazione del numero di strumenti estratti, in quanto nuove estrazioni non consentivano di identificare nuovi strumenti rispetto a quelli collezionati in precedenza.

3.2 Classificazione degli strumenti di *Knowledge Management*

La seconda parte del lavoro consiste nel classificare gli strumenti identificati nel processo del *Knowledge Management*. La classificazione è stata effettuata con l'utilizzo di due metodologie:

- *Top-Down*: assegnare gli strumenti alle quattro fasi del processo di *Knowledge Management*, mediante l'estrazione dei verbi significativi presenti nelle definizioni degli strumenti su Wikipedia. L'estrazione di tali verbi è stata effettuata con l'utilizzo di strumenti di estrazione e manipolazione dei dati. Dall'estrazione è stato creato un dizionario contenente tutti i verbi presenti nelle definizioni e successivamente, sono assegnati alle quattro fasi del processo, in maniera tale da riuscire ad associare gli strumenti di *Knowledge Management* alle fasi di appartenenza.
- *Bottom-Up*: determinare le relazioni esistenti tra gli strumenti, definendo dei cluster di raggruppamento, dove per relazioni si intendono i collegamenti esistenti tra gli *hyperlink* degli strumenti presenti su Wikipedia. Per evidenziare le relazioni è stato creato un grafo in cui i nodi corrispondono agli strumenti e i link rappresentano i collegamenti tra gli strumenti.

⁵ Chiarello, F., Trivelli, L., Bonaccorsi, A., & Fantoni, G. (2018). Extracting and mapping industry 4.0 technologies using wikipedia. *Computers in Industry*, 100(May), 244–257. <https://doi.org/10.1016/j.compind.2018.04.006>

⁶ Hyperlink - Wikipedia. (n.d.). Retrieved February 16, 2021, from <https://en.wikipedia.org/wiki/Hyperlink>.

⁷ <https://cran.r-project.org/web/packages/WikipediaR/index.html>, libreria di R utilizzata per estrarre i dati da Wikipedia.

4 Risultati

Dopo aver definito le metodologie, di seguito vengono mostrati i risultati più significativi ottenuti sia per quanto riguarda la mappatura e identificazione degli strumenti di *Knowledge Management* e sia per classificazione *Top-Down* e *Bottom-Up*.

4.1 Identificazione degli strumenti di *Knowledge Management*

Seguendo i passi definiti nella metodologia, il primo risultato riguarda la ricerca degli strumenti utilizzati nelle *business functions*. Per effettuare l'estrazione degli *hyperlink*, è stata creata una lista contenente il nome dello strumento ed il rispettivo link Wikipedia.

Le iterazioni del processo di estrazione sono state cinque. Il procedimento prevede di inserire in input la lista iniziale di strumenti, dagli *hyperlink* estratti sono stati identificati i nuovi strumenti, dati dal numero totale di estrazioni, eliminando le entità già presenti nella lista precedente, e il numero totale degli strumenti identificati. Partendo da una lista di 96 strumenti, alla fine del processo di estrazione, sono stati identificati 296 strumenti di *Knowledge management*. Successivamente, sono state rimosse tutte le entità con lo stesso link Wikipedia, definite come sinonimi, ottenendo una lista di 245 strumenti di *Knowledge Management*.

In Tab. 4.1 viene riportato un esempio delle entità identificate come sinonimi.

Tabella 4.1: Esempio di identificazione sinonimi

Knowledge Management Tools	Link Wikipedia	
PEST analysis	https://en.wikipedia.org/wiki/PEST_analysis	
PESTLE	https://en.wikipedia.org/wiki/PEST_analysis	
Knowledge Management Tools	Link Wikipedia	Sinonimi
PEST analysis	https://en.wikipedia.org/wiki/PEST_analysis	PESTLE

4.2 Classificazione degli strumenti di *Knowledge Management*

4.2.1 Metodo *Top-Down*

Con la creazione del dizionario di verbi, sono stati assegnati gli strumenti andando ad analizzare la definizione su Wikipedia ed è stato verificato che i verbi contenuti nel dizionario fossero presenti, e tramite strumenti di estrazione automatica del testo, è stata effettuata la classificazione. Di seguito viene riportato un esempio in riferimento alla creazione del dizionario:

Fault tree analysis⁸: *Fault tree analysis is a top-down, deductive failure analysis in which an undesired state of a system is analyzed using Boolean logic to combine a series of lower-level events. This analysis method is mainly used in safety engineering and reliability engineering to understand how systems can fail, to identify the best ways to reduce risk and to determine event rates of a safety accident or a particular system level failure.*

Questa definizione contiene i verbi significativi *to understand how e to identify*, assegnati alle fasi rispettivamente di generazione della conoscenza e identificazione della conoscenza.

In Tab. 4.2 viene riportato un esempio che mostra come sono stati classificati gli strumenti ed è stato utilizzato un metodo di assegnazione binario, indicando con 1 se lo strumento appartiene alla fase e 0 se non appartiene alla fase.

La tabella completa contenente la classificazione *Top-Down*, è consultabile al seguente link:

https://github.com/FabioMalluzzo/Top-Down_Classification.git

Tabella 4.2: Esempio di Classificazione degli strumenti di Knowledge Management

Tools	Link Wikipedia	Creation	Identificaiton	Organization	Application
SWOT analysis	https://en.wikipedia.org/wiki/SWOT_analysis	0	0	0	1
PEST analysis	https://en.wikipedia.org/wiki/PEST_analysis	1	1	0	0

Come si nota in Tab. 4.2, gli strumenti possono appartenere anche a più fasi del processo.

Questo risultato è dato dal fatto che, in un processo aziendale vengono utilizzati strumenti per svolgere più attività e in momenti temporali diversi.

Per avere una valutazione della bontà del metodo, è stato effettuato un confronto tra strumenti appartenenti alle intersezioni tra le fasi del processo e strumenti appartenenti a fasi distaccate.

⁸ Fault tree analysis - Wikipedia. (n.d.). Retrieved April 6, 2021, from https://en.wikipedia.org/wiki/Fault_tree_analysis

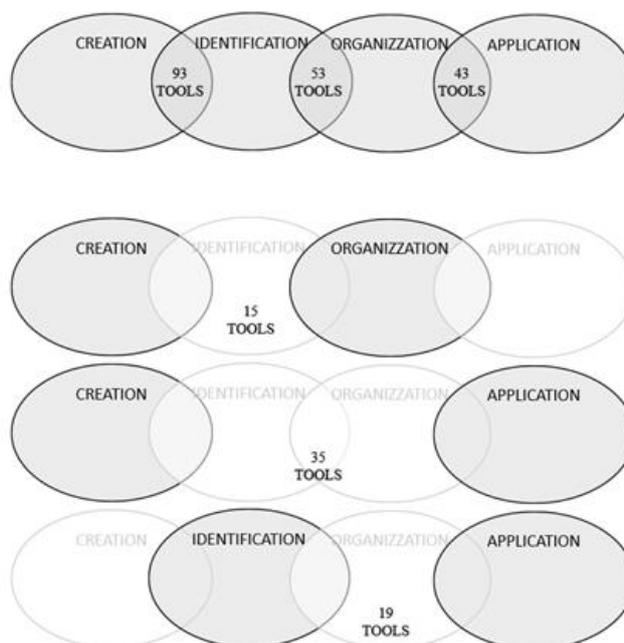
In Tab. 4.3 è riportato il riepilogo degli strumenti assegnati alle quattro fasi.

Tabella 4.3: Riepilogo Classificazione Top-Down

Knowledge Creation	Knowledge Identification	Knowledge Organization	Knowledge Application
181 Tools	148 Tools	91 Tools	130 Tools

Figura 4.1: Strumenti appartenenti alle intersezioni tra Fasi

Dalla Fig. 4.1 si evince come gli strumenti appartenenti alle intersezioni tra le fasi sono maggiori rispetto agli strumenti appartenenti alle fasi distaccate. Questo risultato è dato dal fatto che, uno strumento appartenente a più fasi del processo svolge un ruolo in attività consecutive, in accordo con la metodologia nel determinare la bontà del metodo di classificazione utilizzato.



4.2.2 Metodo Bottom-Up

Il primo passo è stato quello di estrarre gli *hyperlink* presenti nelle pagine Wikipedia degli strumenti e determinare tutte le coppie tra strumenti.

In Tab. 4.4 viene riportato un esempio della tabella contenente le coppie estratte.

Tabella 4.4: Esempio estrazione coppie strumenti e *Hyperlink*

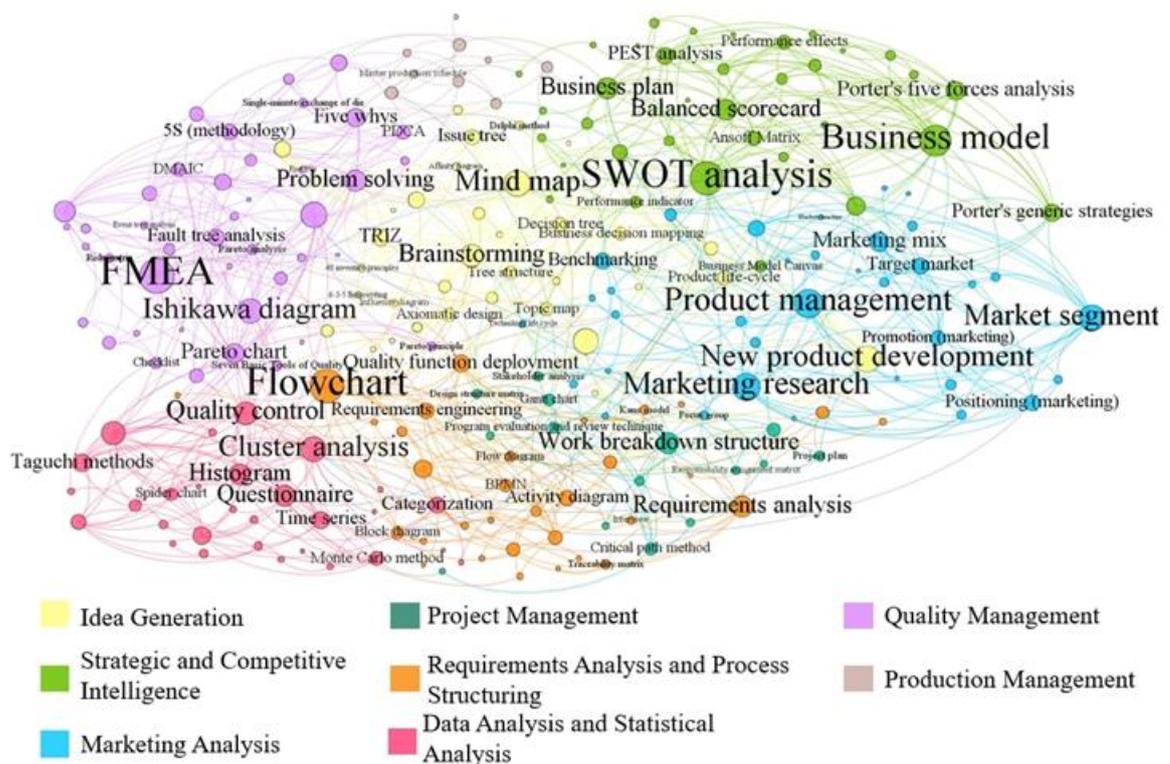
Knowledge Management Tool	Hyperlink	Knowledge Management Tool	Hyperlink
SWOT analysis	Ansoff Matrix	PEST analysis	Ansoff Matrix
SWOT analysis	Balanced scorecard	PEST analysis	Balanced scorecard
SWOT analysis	Mind map	PEST analysis	Competitor analysis
SWOT analysis	PEST analysis	PEST analysis	Mind map
SWOT analysis	Performance effects	PEST analysis	Performance effects
[...]	[...]	[...]	[...]

Successivamente, è stato creato un grafo, in maniera tale da avere una visualizzazione grafica delle relazioni e determinare i cluster di raggruppamento, seguendo l'algoritmo di *Louvain*⁹.

Per la creazione del grafo sono stati definiti i nodi, corrispondenti agli strumenti, e i link, ovvero i collegamenti tra le coppie. Utilizzando l'algoritmo di *Louvain*, sono stati identificati otto cluster, in cui ogni cluster contiene strumenti che sono in relazione tra di loro in riferimento al numero di volte che sono stati menzionati tra gli *hyperlink*. I nomi assegnati ai cluster sono stati definiti tenendo conto delle definizioni degli strumenti contenuti al loro interno.

Come si nota in Fig. 4.2, al centro è presente il cluster *Idea Generation*, contenente strumenti utilizzati per generare conoscenza e quindi si evince il fatto che riesce a confinare con tutti i cluster identificati.

Figura 4.2: Cluster degli strumenti di Knowledge Management



⁹ Louvain method - Wikipedia. (n.d.). Retrieved April 18, 2021, from https://en.wikipedia.org/wiki/Louvain_method

Seguendo, si ha il cluster *Strategic and Competitive Intelligence* e procedendo in senso orario, quindi strumenti strategici per poi passare ad un'analisi del mercato con il cluster *Marketing Analysis*, successivamente si ha la fase di progettazione specifica e le attività di *project management* e analisi dei requisiti.

Infine, si hanno le attività di analisi dei dati, nel cluster *Data Analysis and Statistical Analysis*, per arrivare alle attività di gestione della qualità, con il cluster *Quality Management*.

4.3 Confronto metodi *Top-Down* e *Bottom-up*

La parte conclusiva di questa analisi riguarda il confronto tra la classificazione *Top-Down* e *Bottom-Up*, e determinare se esiste una correlazione tra i due metodi di classificazione.

Per effettuare il confronto è stata creata una tabella strutturata in maniera tale da evidenziare le percentuali degli strumenti presenti in ogni cluster, identificati nel processo di *Knowledge Management*.

Tabella 4.5: Confronto metodi *Top-down* e *Bottom-up*

CLUSTER	KNOWLEDGE MANAGEMENT PROCESS			
	CREATION	IDENTIFICATION	ORGANIZATION	APPLICATION
Idea Generation	30%	25%	16%	29%
Strategic and Competitive Intelligence	38%	31%	8%	23%
Marketing Analysis	34%	36%	15%	15%
Project Management	31%	31%	7%	31%
Requirements Analysis and Process Structuring	41%	19%	14%	26%
Data Analysis and Statistical Analysis	37%	25%	18%	20%
Quality Management	25%	24%	24%	27%
Production Management	37%	42%	0	21%

Dalla Tab. 4.5 alcune considerazioni possono essere fatte per i cluster "*Strategic and Competitive Intelligence*", "*Project Management e Requirements Analysis and Process Structuring*", con un riscontro tra i due metodi, riferito al fatto che la maggiore percentuale di strumenti facenti parte di questi cluster appartengono alle fasi di creazione e identificazione della conoscenza, in accordo con le definizioni dei cluster, rispetto ad esempio al cluster "*Production Management*" con una percentuale nulla per quanto gli strumenti appartenenti alla fase di organizzazione della conoscenza, quindi non si ha un riscontro tra i due metodi.

È possibile quindi vedere come le due classificazioni forniscono due informazioni differenti. La classificazione *Top-Down* identifica quali sono le fasi di gestione della conoscenza in cui lo strumento è impiegato.

La classificazione *Bottom-up* fornisce una mappatura degli strumenti in base alle funzioni di *business*.

5 Conclusioni e Sviluppi futuri

L'obiettivo del lavoro di tesi è stato quello di creare un *framework* di strumenti o metodi di gestione della conoscenza in maniera tale da riuscire ad effettuare una classificazione degli strumenti all'interno del processo di *Knowledge Management*. In particolare, si ritiene che lo strumento fornito possa fornire un aiuto nell'adozione di un determinato metodo in certe fasi, per raggiungere un vantaggio competitivo all'interno della propria organizzazione. Il risultato è particolarmente utile anche per approcciare l'insegnamento di tali strumenti, in quanto fornisce una chiara divisione concettuale tra diverse tipologie di strumenti. Infine, il grafo proposto può essere visto come un manuale apio di strumenti diversi, che può essere utilizzato da chiunque utilizzi questi strumenti nel proprio lavoro, per cercare strumenti simili e complementari.

La futura analisi potrebbe essere riferita alla ricerca degli strumenti ampliando le fonti di ricerca, per avere sia una ulteriore espansione del *framework* di strumenti, che una diversa definizione degli strumenti.