DEMB Working Paper Series

N. 74

Promozione del sistema dei poli di innovazione nello spazio web: analisi dei contenuti e delle reti di relazioni virtuali

Matteo Di Cristofaro*, Valentina Fiordelmondo**,
Stefano Ghinoi**, Margherita Russo***

December 2015

* PhD in Corpus Linguistics, Lancaster University
email: deek85@gmail.com

** Progetto Poli.in

*** University of Modena and Reggio Emilia
CAPP, Centre for the Analysis of Public Policies
Address: Viale Berengario 51, 41121 Modena, Italy
email: margherita.russo @unimore.it

ISSN: 2281-440X online

Dipartimento di Economia Marco Biagi
Università degli studi di Modena e Reggio Emilia
Via Berengario 51 | 41121 Modena
tel. 059 2056711 | fax. 059 2056937
info.economia@unimore.it | www.economia.unimore.it

Promozione del sistema dei poli di innovazione nello spazio web: analisi dei contenuti e delle reti di relazioni virtuali

Promotion of the system of innovation poles in the web space: analysis of content and networks of virtual relationships

Matteo Di Cristofaro[*], Valentina Fiordelmondo[§],
Stefano Ghinoi[§], Margherita Russo[^]

[*] PhD in Corpus Linguistics, Lancaster University, deek85@gmail.com

[^] Dipartimento di Economia Marco Biagi e CAPP, Università di Modena e Reggio Emilia, margherita.russo@unimore.it

[§] Progetto Poli.in

Abstract

Nel progetto di analisi e modellizzazione del sistema regionale dei poli di innovazione in Toscana (www.poliinnovazione.unimore.it), abbiamo esaminato in modo sistematico le informazioni disponibili nei siti web dei poli con due obiettivi: (i) analizzare la varietà di linguaggio e di contenuti che caratterizzano i poli nella loro attività on line; (ii) verificare in che misura i poli facciano riferimento agli stessi enti, imprese, organizzazioni, progetti e, tra questi, alle organizzazioni/o attività direttamente collegate al polo (come ad esempio i soggetti gestori dei poli laboratori, gli incubatori, gli aderenti). Questo articolo presenta nella prima parte l'analisi linguistica dei siti web prodotti dai poli di innovazione ed ha lo scopo di fornire un insieme di variabili per la valutazione delle loro strategie di promozione. Il metro di valutazione principale è l'individuazione della fonte dei materiali pubblicati online dai singoli poli (materiali originali vs. materiali prodotti da terzi), che è stato usato come indicatore del livello di partecipazione di ogni singolo polo nelle strategie di promozione delle proprie attività. Si è quindi cercato di quantificare il livello di coinvolgimento che ogni polo ha espresso – nel proprio sito web – attraverso la pubblicazione di testi. L'analisi è stata condotta utilizzando la metodologia di linguistica dei corpora al fine di svolgere un'indagine sia quantitativa sia qualitativa (simile alla metodologia nota come CADS), ed è stata effettuata prendendo in esame 24 termini relativi ai concetti di innovazione e sviluppo, all'interno dei contenuti web pubblicati dai poli di innovazione (azienda, aziende, centri, collaborazione, conoscenze, finanziamenti, gestione, gestore, impresa, imprese, industriale, innovazione, poli, polo, processi, progetti, progetto, ricerca, servizi, servizio, sviluppo, tecnologia, tecnologie, territorio). Attraverso la valutazione dell'utilizzo dei 24 termini all'interno dei testi web, l'analisi ha evidenziato come le nozioni di innovazione e sviluppo vengano costruite dai singoli poli in termini sia di significato, sia di connotazione.
Oltre all'analisi linguistica, l'analisi dei testi web ha consentito anche di individuare in maniera puntuale i domini principali e i link completi citati nei siti web dei poli. Adottando una prospettiva di analisi delle reti di relazioni, con queste informazioni sulle reti virtuali abbiamo analizzato due questioni rilevanti per la modellizzazione del sistema dei poli di innovazione: attraverso le citazioni dei domini principali mettiamo in evidenza in che misura i siti web i poli hanno comunicato le connessioni tra di loro e con i soggetti attivi nel trasferimento tecnologico; attraverso l'analisi dei link completi presenti nei siti web dei singoli poli possiamo individuare in che misura i poli fanno riferimento allo stesso spazio di informazioni.
La presentazione è strutturata come segue: nella sezione 1 vengono presentati i dati, gli strumenti e la metodologia utilizzata nell'analisi linguistica; nella sezione 2 viene presentata l'analisi linguistica dei 24 termini selezionati ; nella sezione 3 sono analizzate le reti di relazioni virtuali, la sezione 4 conclude. L'appendice 1 presenta i dettagli tecnici sulla raccolta e successiva pulizia dei dati; l'appendice 2 contiene la lista dei principali link completi presenti all'interno dei siti dei poli presi in esame.

In analysing and modelling the Tuscany regional system of innovation poles (www.poliinnovazione.unimore.it), we examine systematically available information on websites with two objectives: (i) to analyze the variety of language and content that characterize the poles in their online activities; (ii) examine the extent poles refer to the same institutions, enterprises, organizations, projects, and among these, organizations / or activities directly related to the (such as the companies managing the poles laboratories, incubators, the adherents). This paper offers a linguistic analysis of the websites produced by innovation poles in Tuscany, in order to provide a set of variables for the evaluation of their promotion strategies. The focus of the analysis is the evaluation of original vs. reported information, which is here used as indicator of the innovation poles' engagement in the promotion of their activities. The purpose is to quantify the level of engagement that each single network has built – on its website – through the publication of texts.

Through Corpus Linguistics methodology, and using both a quantitative and a qualitative approach (similar to the one proposed by the CADS framework), this papers looks at how the innovation poles taken into account have used a set of 24 terms related to the notions of innovation and promotion (business, businesses, centers, collaboration, knowledge, finance, management, manager, business, business, industrial, innovation, pole, poles, processes, projects, design, research, service, services, development, technology , technologies, territory). By looking at whether these terms appear in original or reported (produced by third-parties) texts, and at how their meanings and connotations are constructed in the former cases, this paper identifies how active have the innovation poles been in promoting their work.

In addition to linguistic analysis, text analysis of web pages also allowed to identify main domains (urls) and complete links mentioned on the websites of poles. Adopting a perspective of social network analysis, with information on these virtual networks we analysed two issues relevant to modelling the system of innovation poles: through the quotes of major domains, we highlight the extent to which the web sites have reported the poles connections with each other and with those involved in technology transfer; through the analysis of complete links present on the websites of individual poles we can identify to what extent the poles refer to the same information space.

The presentation is structured as follows: section 1 presents the data, tools and methodology used in linguistic analysis; section 2 presents linguistic analysis of the 24 selected terms; in section 3 virtual networks are analysed; section 4 concludes. Appendix 1 presents the technical details on the collection and subsequent data cleansing; Appendix 2 contains the list of the top link complete inside of the poles of the sites surveyed.

Key words: linguistic analysis; web communication; innovation poles; regional innovation policies; network analysis

JEL codes: R10 General Regional Economics; O25 Industrial Policy; Y8 Related Disciplines; C88 Other Computer Software

Credits

Il progetto di ricerca "Poli.in_analisi e modellizzazione dei poli di innovazione in Toscana" è frutto dell'accordo interistituzionale tra la Regione Toscana e l'Università degli studi di Modena e Reggio Emilia – Dipartimento di Economia Marco Biagi. Oggetto dell'accordo è la collaborazione finalizzata al supporto in attività di valutazione e modellizzazione delle dinamiche di trasferimento tecnologico in seno ai Poli di innovazione approvati con DD n. 2859/2011.

Ogni responsabilità su quanto riportato nel testo resta da attribuire agli autori. Il progetto, condotto dagli autori in stretta collaborazione, è stato sviluppato da Matteo di Cristofaro per la parte di analisi linguistica dei siti web e per la messa a punto delle metodologie di *text mining*; Valentina Fiordelmondo ha curato in particolare la raccolta dei dati preliminari per l'analisi dei siti web e la verifica dei risultati necessari per l'analisi delle reti; Stefano Ghinoi ha elaborato la sezione relativa ai collegamenti tra siti web. Margherita Russo, responsabile del progetto Poli.in, ha ideato e coordinato il progetto, del quale ha curato la stesura finale.

Premessa

Nel periodo 2011-2013 la Regione Toscana ha sostenuto la creazione di 12 poli di innovazione con l'obiettivo di sostenere un'offerta di servizi per il trasferimento tecnologico, in diverse specializzazioni e territori della regione. Questo contributo, sviluppato nell'ambito del progetto di ricerca Poli.in_Analisi e modellizzazione dei poli di innovazione in Toscana, utilizza l'analisi linguistica dei siti web per individuare le tematiche che caratterizzano i poli di innovazione e per delineare lo spazio di relazioni virtuali che emerge dalla citazioni dei domini contenuti nelle pagine web dei poli[1].

Tra le fonti di informazione consultate nella primissima fase della ricerca Poli.in vi era la documentazione prodotta dai poli nei loro siti web a un anno di avvio della loro attività dei poli[2]. A conclusione dell'attività dei poli abbiamo riesaminato in modo sistematico le informazioni disponibili nei siti web con due obiettivi: (i) analizzare la varietà di linguaggio e di contenuti che caratterizzano i poli nella loro attività on line; (ii) verificare in che misura i poli facciano riferimento agli stessi enti, imprese, organizzazioni, progetti e, tra questi, alle organizzazioni/o attività direttamente collegate al polo (come ad esempio i soggetti gestori, i laboratori, gli incubatori, gli aderenti). Un terzo obiettivo, quello di estrarre dai siti web informazioni qualitative sull'attività di animazione svolta dai poli, non ha prodotto risultati significativi e non viene quindi preso in esame in questo documento.

Nella sezione 1 vengono presentati i dati, gli strumenti e la metodologia utilizzata nell'analisi linguistica; nella sezione 2 viene presentata l'analisi linguistica dei 24 termini selezionati ; nella sezione 3 sono analizzate le reti di relazioni virtuali, la sezione 4 conclude. L'appendice 1 presenta i dettagli tecnici sulla raccolta e successiva pulizia dei dati; l'appendice 2 contiene la lista dei principali link completi presenti all'interno dei siti dei poli presi in esame.

1 Analisi dei siti internet dei poli

Rispetto alla prima analisi dei siti web dei poli di innovazione condotta nel 2014, nell'analisi che segue, condotta alla conclusione del triennio di operatività dei poli, constatiamo che alcuni siti web non sono più accessibili[3]. Nell'analisi che segue faremo quindi riferimento ai siti web di tutti i poli di innovazione, a meno del polo Pietre, che non era online al momento della raccolta dati. Per il polo Vita si è fatto riferimento a sito web del nuovo distretto Scienze della Vita.

Al fine di analizzare il linguaggio utilizzato dagli 11 poli nei rispettivi siti internet, sono stati innanzitutto scaricati i soli file .html da ogni sito, tramite l'utilizzo di *httrack*. I file sono poi stati puliti togliendo il materiale non utile ai fini dell'analisi linguistica, come ad esempio il codice html, disclaimer, ecc... Per maggiori informazioni sulla raccolta e sulla pulizia si veda il documento "Metodologia". Nota: il corpus www.polis-toscana.it utilizzato nell'analisi – e pertanto tutti i dati presentati in questa analisi – contiene 13.531 pagine web, un quinto del totale delle pagine presenti sul sito del polo (67.656). È stato infatti

[1] In un lavoro complementare viene sviluppata l'analisi dei testi prodotti dai poli di innovazione nel corso della loro attività, si veda P. Pavone, V. Fiordelmondo, M. Russo, 2015, Le parole dei poli di innovazione: dall'analisi dei linguaggi alla modellizzazione del sistema regionale di innovazione in Toscana, DEMB Working Paper Series n.

[2] L'analisi era stata presentata in occasione della prima presentazione pubblica della ricerca (il 7 febbraio 2014, si veda http://www.poliinnovazione.unimore.it/?p=521).

[3] Polo Pietre: il sito www.polopietretoscane.it non è più attivo. Polo Vita: per questo polo di innovazione non è mai esistito un vero e proprio sito web, ma il capofila Fondazione Toscana Life Science, aveva previsto una pagina descrittiva all'interno della sezione “progetti” del suo sito internet, concentrando invece tutta l'informazione sul sito del distretto scienza della vita, diventato oggi il sito web di riferimento e sul quale si sono sviluppate le analisi di questo capito.

necessario ridurre il numero di file da analizzare, in modo da rendere possibile il caricamento del corpus da parte di AntConc. La selezione dei file da utilizzare come corpus è avvenuta in maniera casuale (*simple random sampling*), attraverso l'utilizzo del programma *shuf*, specifico per la selezione casuale di elementi. Ciò al fine di mantenere il più possibile intatta la *representativeness* del corpus, ovvero quanto il corpus è "rappresentativo" dell'insieme dei testi presi in considerazioni. Non sono state necessarie operazioni per garantire la *balance* dei corpora (ovvero il rapporto tra i differenti tipi di testi dell'insieme preso in considerazione) in quanto le singole pagine di ogni sito – polistoscana.it incluso – sono state considerate come appartenenti ad un'unica tipologia, i.e. quella della pagina web appartenente al sito internet di un polo industriale.

La Tabella 1 descrive sinteticamente i corpora dei siti web presi in esame.

Tabella 1 Presentazione dei corpora dei siti web degli 11 poli analizzati

	Tokens numero delle parole contenute nel corpus	**Types** numero di termini unici che appaiono nel corpus	**TTR** rapporto tra il numero totale di parole e il numero di termini unici
innopaper.lucense.it	9 921	2 128	21,45%
optoscana.net	26 776	2 470	9,22%
polocento.it	64 399	4 375	6,79%
www.distrettoict-robotica.it	94 497	7 749	8,20%
www.nanoxm.it	46 680	5 027	10,77%
www.otir2020.it	9 329	1 130	12,11%
www.polis-toscana.it	1.588.739	1 380	0,09%
www.polo12.it	119.416	2 930	2,45%
www.polopenta.it	96 270	2 516	2,61%
www.polopierre.it	16.736	2 376	14,20%
www.scienzedellavita.it	627.217	7 308	1,17%

Fonte: ns. elaborazioni su dati tratti dai siti web. Impostazioni AntConc 3.2.4u (Linux): Encoding: UTF8; Token: Letter

Tokens indica il numero delle parole contenute nel corpus; *Types* indica il numero di termini unici che appaiono nel corpus; *TTR* indica il *Type-token ratio*, ovvero il rapporto tra il numero totale di parole e il numero di termini unici, che è indice di ricchezza lessicale[4]. Più il rapporto è basso, più i testi tendono a contenere un maggior numero di parole uguali (i.e. ad essere composti dalle stesse parole). Il *TTR* è pertanto strettamente correlato alla dimensione del corpus, siccome una maggiore quantità di testo implica una maggiore frequenza di parole grammaticali/funzionali (preposizioni, congiunzioni, ecc...), la quale a sua volta comporta una diminuzione del valore di TTR[5]. Al fine di limitare la distorsione dei risultati è pertanto prassi nella linguistica dei corpora l'utilizzo del *Standardized Type-Token Ratio* (*STTR*). Questo calcolo tiene conto della dimensione del corpus, e si effettua dividendo prima il corpus in insiemi di parole (solitamente 1.000 o 2.000) e successivamente calcolando il TTR per ognuno degli insiemi. I valori di TTR dei singoli insiemi vengono poi raccolti, e la media dei valori rappresenta il STTR. Nell'analisi qui presentata però si è deciso di non standardizzare i valori di TTR – di non usare pertanto il STTR – ma di utilizzare i soli valori di TTR. La motivazione risiede nello scopo di questa ricerca, che è stabilire quanto ogni singolo polo contribuisca con materiale proprio al sito del polo stesso.

[4] Baker, Hardie, McEnery, *A Glossary of Corpus Linguistics*, 2006:162--163, Edinburgh University Press

[5] Cfr. Baker, Hardie, McEnery 2006:150--151,163

Si prendano ad esempio i corpora dei poli *innopaper* e *polo12*: essi presentano valori TTR rispettivamente di 21.45% e 2.45%. I valori indicano che innopaper presenta un maggiore ricchezza lessicale di polo12. Questo risultato è facilmente prevedibile siccome i dati di polo12 contengono 119.416 tokens, mentre quelli di innopaper 9.921; è dunque statisticamente più probabile che i dati di polo12 contengano un numero maggiore di parole grammaticali/funzionali rispetto ad innopaper. Se si controllano però i valori standardizzati di TTR – ovvero STTR - si nota come innopaper presenti un valore di 46.45%, mentre polo12 di 47.38%; lo STTR sembrerebbe dunque suggerire che i dati polo12 siano leggermente più ricchi a livello lessicale. L'analisi eseguita sui corpora dei siti ha però mostrato come i testi contenuti nel corpus di polo12 siano prevalentemente provenienti da siti terzi, e come molti testi siano ripetuti molteplici volte all'interno dei dati; questo non accade invece con innopaper. Siccome lo scopo principale dell'analisi è capire il livello di "coinvolgimento" del polo nella "costruzione" del proprio sito, la metodologia utilizzata ha dato un maggior peso ai testi scritti dai poli che a quelli riportati da siti terzi; e l'analisi ha evidenziato come la maggior parte dei testi ripetuti molteplici volte provenga da siti terzi. Appare dunque evidente come l'utilizzo del TTR – per quanto teoricamente non congruo alla comparazione di corpora di differenti dimensioni – sia in questo caso preferibile al STTR. TTR mette infatti in evidenza la ricchezza lessicale di tutti i testi presi nel loro insieme, mentre STTR divide il testo in gruppi da 1000 parole ognuno, limitando dunque la possibilità che le molteplici copie di uno stesso testo vengano prese in considerazione nel calcolo del TTR. La spiegazione di questa preferenza può essere esemplicata nel seguente modo: si prendano due insiemi (A e B) di 20 articoli ciascuno; gli articoli dell'insieme A sono tutti diversi l'uno dall'altro e sono formati da 5000 parole ciascuno, quelli dell'insieme B sono in parte ripetuti e sono formati da 10000 parole ciascuno. Calcolando il TTR dei due insiemi, il valore per l'insieme A sarà necessariamente molto più alto del valore dell'insieme B, siccome una parte dei testi di quest'ultimo insieme è formato da ripetizioni di un unico articolo. Il calcolo del STTR potrebbe invece presentare un valore più alto per l'insieme B, siccome ogni articolo verrebbe diviso in 10 differenti gruppi da 1000 parole ognuno, e il TTR di ognuno di questi verrebbe poi utilizzato per calcolare la media di tutti i TTR – ovvero STTR.

I valori di STTR rendono pertanto più difficile l'individuazione di porzioni di testo ripetute, ed essendo quest'ultimo un parametro fondamentale per l'analisi linguistica dei corpora dei poli, si è optato per l'utilizzo del TTR al fine di includere nel valore non solo la maggiore frequenza di parole grammaticali/funzionali, ma anche le ripetizioni di uguali porzioni di testo.

Una volta creati gli 11 corpora – i cui dettagli sono visibili nella Tabella 1 – si è provveduto ad estrarre, per ognuno di essi, una "*word list*" con l'elenco di tutti i termini (types) presenti nel corpus, ordinati per frequenza e corredati della frequenza assoluta all'interno dei dati. Le 11 "word list" sono state utilizzate per identificare quei termini che vengono utilizzati in tutti gli 11 corpora, al fine di condurre un'analisi comparativa di specifici termini. Una volta creato l'elenco dei termini più frequenti negli 11 corpora, sono stati selezionati 24 termini di rilievo per l'argomento di ricerca, per i quali si è analizzato in che modo vengono utilizzati all'interno di ogni singolo sito. La metodologia utilizzata sia per l'estrazione delle *word list* sia per l'analisi linguistica dei termini – attraverso l'uso di *concordances* e *collocations* – segue gli standard della linguistica dei corpora (per maggiori dettagli si veda McEnery, Hardie, *Corpus Linguistics: Method, Theory and Practice*, 2012 CUP).

La metodologia è stata poi adattata allo scopo di questa elaborazione cioè quello di cogliere il contesto in cui i termini selezionati vengono utilizzati. In particolare, la distinzione principale è data dalla provenienza del testo, ovvero: il testo è stato prodotto dal polo che l'ha pubblicato, o è preso da un sito terzo? L'obiettivo di questa analisi è quello di verificare il livello di coinvolgimento da parte dei singoli poli rispetto ai contenuti dei siti, e di valutare l'attività dei poli in quanto generatori di connessioni di informazioni rilevanti

per chi consulta il sito web del polo. Il metro di valutazione riguarda pertanto l'analisi dei termini selezionati all'interno delle tipologie dei testi.

Questa distinzione risulta essere molto importante se consideriamo il sito internet come una vetrina attraverso la quale il polo promuove le proprie attività, costruisce la propria immagine/identità ed informa l'utente. Risulta pertanto cruciale il ruolo che ha la lingua nel veicolare il messaggio. Un'analisi di come vengono usate specifiche parole inerenti l'attività dei poli e l'ambiente in cui operano può fornire indicazioni utili per un'analisi qualitativa dell'azione intrapresa dai poli di innovazione nella comunicazione della propria attività e dell'ambiente in cui opera.

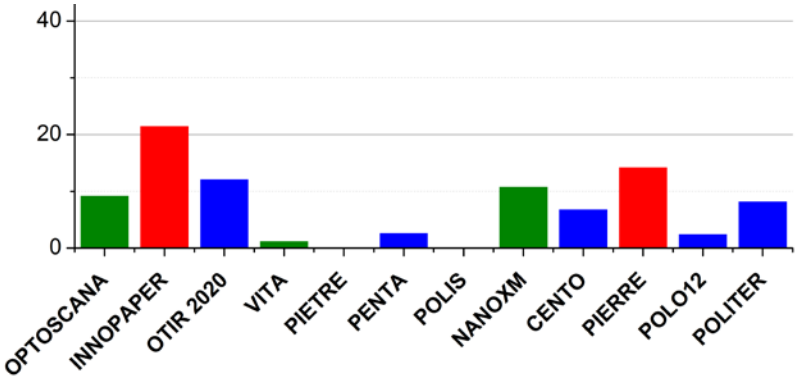
L'analisi ha mostrato che la maggior parte dei poli riporta testi (notizie, bandi, ecc...) tratti da altri siti, in cui i termini presi in analisi appaiono come parte di titoli o descrizioni, principalmente all'interno di bandi o descrizioni di eventi. In questo caso dunque i termini non vengono direttamente prodotti dai poli, che non costruiscono pertanto un significato specifico del termine, né lo caratterizzano/connotano secondo le proprie esigenze.

Ad esempio, il termine *gestore* appare quasi esclusivamente all'interno della costruzione *soggetto gestore* all'interno degli 11 siti, e viene quasi esclusivamente utilizzato come appellativo. In questo caso pertanto *gestore* non è una parola che viene “attivamente” utilizzata ad esempio per promuovere o descrivere le caratteristiche (attività, prodotti, etc...) dell'entità a cui si riferisce, ma indica semplicemente una caratteristica (il ruolo di *soggetto gestore*) del polo che viene citato.

Un ulteriore esempio è rappresentato dai contenuti presenti nel sito di polis-toscana.it, il sito con la maggior quantità di contenuti. Un'analisi dei termini e dei contenuti ha mostrato che la totalità dei testi sono presi da siti terzi, all'interno di pagine in cui polis-toscana.it riporta semplicemente documenti prodotti da altri siti (non bisogna però dimenticare che i dati relativi a polis-toscana.it rappresentano un quinto del totale dei file del sito; v. nota prima della tabella n.22). In questo caso dunque la 'costruzione' di concetti attorno ai termini da parte di polis-toscana.it è praticamente nulla, ma è elevatissima la generatività potenziale di connessioni con altri, domini che arricchiscono quindi la comunicazione offerta agli utenti del sito.

Risulta quindi possibile creare una sorta di continuum tra due estremi : da un lato i poli che utilizzano il proprio sito internet per informare gli utenti delle proprie novità, e per promuoversi. Dall'altro quei poli i cui siti fungono da “aggregatori” di notizie riguardanti il “mondo dei poli”, e più in generale il “mondo del trasferimento tecnologico”, senza però focalizzarsi sul polo proprietario del sito. L'esempio soprariportato di polis-toscana è forse il più rappresentativo di quest'ultima tipologia di siti: il sito polis-toscana è un enorme aggregatore di eventi, bandi, notizie, report, relativi ad una serie di poli ed aziende legate al trasferimento tecnologico. Tra i poli che più sfruttano il sito per promuoversi ed informare gli utenti riguardo alle proprie attività c'è sicuramente innopaper, che utilizza quasi tutti i termini analizzati in modo diretto (utilizzandoli all'interno di testi creati dal polo stesso). È interessante notare come i due siti, innopaper e polis-toscana (gli estremi del continuum), siano anche i due siti che presentano maggiori differenze nei valori di Type-Token Ratio, rispettivamente 21,45% e 0,09% (Figura 1). Ciò non è inaspettato: il sito innopaper è ricco di contenuti originali e quasi mai ripetuti, il che porta ad un'alta varietà di vocabolario. Polis-toscana, riportando contenuti non originali (spesso molto brevi) e frequentemente ripetuti, presenta un vocabolario molto più limitato.

Figura 1 Indice di varietà lessicale(Type-token ratio: rapporto tra il numero totale di parole e il numero di termini unici), per polo



Fonte: ns. elaborazione su dati tratti dai siti web dei poli di innovazione; estrazione maggio 2015

Un ultimo dettaglio riguardo all'analisi: bisogna considerare che la natura di questa analisi è principalmente qualitativa, piuttosto che quantitativa; si da quindi più importanza al valore delle singole occorrenze e utilizzi dei termini, invece che alla frequenza con cui essi appaiono nei dati. Per esemplificare il motivo di questa scelta, si prenda in considerazione il termine *innovazione*: esso appare principalmente – e con un alta frequenza – all'interno delle costruzioni *poli di innovazione* e *polo di innovazione*. Ciononostante, le occorrenze di queste costruzioni hanno poca importanza nella costruzione del significato del termine inteso come «L'atto, l'opera di innovare, cioè di introdurre nuovi sistemi, nuovi ordinamenti, nuovi metodi di produzione» (http://www.treccani.it/vocabolario/ricerca/innovazione/). Le occorrenze più significative, ovvero quelle in cui *innovazione* è utilizzato per esprimere in maniera esplicita e diretta il significato riportato sopra, sono molto poco frequenti, ma estremamente importanti per poter capire in che modo i poli informano e si promuovono. La scarsa frequenza di quest'ultimo significato, e l'alta frequenza delle costruzioni *polo/i di innovazione*, è facilmente comprensibile: il significato più concreto è usato in testi in cui – ad esempio – il polo descrive un servizio o un progetto, mentre le costruzioni appaiono – ad esempio – in ogni pagina in cui appare il nome del polo. Mentre la descrizione di un servizio/progetto è limitata ad una singola pagina, la dicitura *polo d'innovazione* è riportata più volte all'interno di quasi tutte le pagine del sito.

I termini analizzati sono stati scelti in base alla rilevanza con la tematica della ricerca. Di seguito i 24 termini identificati: azienda, aziende, centri, collaborazione, conoscenze, finanziamenti, gestione, gestore, impresa, imprese, industriale, innovazione, poli, polo, processi, progetti, progetto, ricerca, servizi, servizio, sviluppo, tecnologia, tecnologie, territorio.

2 Analisi dei 24 termini selezionati

Note preliminari

Dall'analisi sono stati quasi del tutto esclusi i corpora di alcuni siti (es. polis-toscana, scienzedellavita) in quanto non presentano alcuna occorrenza diretta dei termini analizzati. Essendo i siti composti di contenuti presi da siti terzi, è praticamente nullo in coinvolgimento dei suddetti poli nella costruzione dei significati dei termini.

La serie di grafici riportati nella Figura 2, posta alla fine di questa sezione, documenta per ciascuno degli undici siti web analizzati, le occorrenze dei 24 termini selezionati (frequenze normalizzate per milione di parole).

AZIENDA

Il termine *azienda* assume tre connotazioni e/o significati diversi negli 11 siti:

i) con funzione di riferimento all'azienda proprietaria del sito;

ii) con riferimento ad un'azienda terza;
iii) senza nessun significato privilegiato, principalmente a causa della bassissima frequenza (<12). optoscana.net e polo12.it usano il termine per riferirsi principalmente a se stessi, mentre distrettoict-robotica.it e polis-toscana.it ne fanno un uso diametralmente opposto, riferendosi ad aziende di terzi (poiché principalmente riportano notizie da siti terzi). A metà si posiziona nanoxm.it che utilizza quasi con lo stesso peso entrambe i significati, mentre per i rimanenti siti il significato rimane più generale quando è presente in bassa frequenza.

AZIENDE

Aziende è riconducibile a due accezioni principali: viene usato per descrivere i servizi/prodotti che il polo offre alle *aziende*; oppure per descrivere le *aziende* che fanno parte del polo.

innopaper.lucense.it, otir2020.it, nanoxm.it utilizzano il termine principalmente per rivolgersi a potenziali clienti, descrivendo i servizi/prodotti offerti rivolti appunto alle "aziende".

polocento.it, distrettoict-robotica.it, polo12.it ne fanno invece un utilizzo quasi esclusivamente diverso, per riferirsi alle *aziende* che fanno parte del polo (eventi a cui esse prendono parte, conseguimento di premi, traguardi, ecc...).

CENTRI

Il termine è spesso riferito a centri di natura universitaria ("centri universitari", "centri servizi-università"), con una minore frequenza di altre 'tipologie' di centri. Solamente 4 poli (innopaper, otir2020, polo12, polopenta) utilizzano "attivamente" il termine per riferirsi alle proprie attività e competenze, mentre i rimanenti poli riportano semplicemente testi di bandi e articoli esterni. Di seguito alcuni esempi tratti dai quattro siti sopracitati:

- "Polo 12" è costituito da una rete di sei centri di formazione,assistenza alle imprese e trasferimento tecnologico regionali (polo12)
- In alcuni casi alcui di questi centri sono fortemente specializzati e si occupano di ricerca esclusivamente in questo settore. (innopaper)
- Le realtà coinvolte sono Centri Servizi e di Ricerca e Aree di ricerca del CNR (otir2020)
- Gestione dei rapporti esterni con centri e laboratori accreditati, veicolazione di conoscenza (otir2020)
- Il polo di innovazione per la nautica e le tecnologie del mare PENTA è lo strumento di congiunzione dei 5 centri servizi toscani che presidiano le imprese sulle varie province di costa (polopenta)
- Oltre alle collaborazioni con università e centri di ricerca il polo per l'innovazione PENTA si è avvalso di consulenze esterne per migliorare e completare il proprio approccio all'impresa. (polopenta)

Nella maggioranza dei casi il termine è usato in maniera semplicemente descrittiva; solo polopenta utilizza il termine all'interno di una costruzione in cui assume un carattere più legato all'attività stessa del polo, "PENTA è lo strumento di congiunzione dei 5 centri servizi toscani"

COLLABORAZIONE

Collaborazione è usato in maniera diretta da 4 poli (innopaper, polocento, nanoxm, polopenta), nonostante solo innopaper lo usi esclusivamente per riferirsi a proprie attività. Nei rimanenti 3 poli il termine appare principalmente all'interno di news e bandi esterni, e solo raramente in testi scritti dai poli stessi.

- INNOPAPER, anche attraverso il personale qualificato di LUCENSE, soggetto gestore del polo di innovazione, e grazie alla collaborazione con esperti universitari, è in grado di erogare numerose tipologie di servizi avanzati e qualificati. (innopaper)

- Il ciclo di incontri "Tecnologie emergenti per la competitività delle imprese", è organizzata da CENTO - Polo di competenza per il settore interni, in collaborazione con la Regione Toscana - POR CReO FESR e Tecnorete Toscana. (polocento)
- * Accordi di collaborazione (nanoxm)
- Le nuove opportunità di finanziamento per le imprese attraverso progetti di Ricerca, Sviluppo ed Innovazione";, organizzato in collaborazione con il Polo per la Fotonica e l'Optoelettronica OPTOSCANA (nanoxm).

CONOSCENZE

Conoscenze è principalmente legato a termini quali "trasferimento", "scambio", "fornire", con una bassa frequenza in tutti gli 11 siti. Solo 3 poli utilizzano il termine rifersi "a se stessi" (innopaper, polo12, polopenta), mentre nanoxm cita, all'interno di un documento di presentazione del polo, un testo esterno. I rimanenti siti riportano testi da siti terzi.

- INNOPAPER ha come obiettivo di stimolare l'attività innovativa incoraggiando l'interazione intensiva, l'uso in comune di installazioni e lo scambio di conoscenze ed esperienze (innopaper, ma poco)
- destinati a stimolare l'attività innovativa incoraggiando l'interazione intensiva, l'uso in comune di installazioni e lo scambio di conoscenze ed esperienze, nonché contribuendo in maniera effettiva al trasferimento di tecnologie (nanoxm, ma è una citazione di un testo che usa nella descrizione di cos'è un polo mentre parla di se stesso!)
- TRASFERIMENTO DI CONOSCENZE dal mondo delle Università e della Ricerca alle imprese. (polo12)
- Il Polo adotta un Programma di attività di trasferimento di conoscenze e competenze tecnologiche e scientifiche di durata triennale (polopenta)

FINANZIAMENTI

Il termine *finanziamenti* presenta una bassa frequenza in tutti gli 11 corpora, e solo 2 poli lo utilizzano in maniera diretta, optoscana e polopenta, per descrivere le proprie attività:

- Per informazioni sul Catalogo vi rimandiamo alla pagina sui finanziamenti regionali QUI. (optoscana)
- I finanaziamenti per Optoscana di dividono in: finanziamenti EUROPEI finanziamenti NAZIONALI finanziamenti REGIONALI (optoscana)
- Sono inoltre stati inseriti nella sezione finanziamenti due bandi, uno nazionale e uno europeo. (optoscana)
- Eccellenza sui temi di: * Organizzazione aziendale * Sviluppo Innovazione * Finanza (possibilità di finanziamenti pubblici e privati) (polopenta)

GESTIONE

Il termine *gestione* è strettamente legato alle attività dei singoli polo: non tutti i poli presi in considerazione operano nel campo della "gestione" (che sia essa "di macchine", "dei processi", ecc...), e pertanto la sua frequenza è bassa. Solo 3 poli (polocento, nanoxm, otir2020) riportano il termine all'interno di descrizioni dei servizi offerti o delle attività svolte, mentre nei rimanenti siti *gestione* è parte di descrizioni non relative al polo il cui sito contiene il testo in questione.

- Il driver di 3S è l'implementazione della gestione elettronica delle macchine al fine di assicurare un'interazione «uomo/macchina» e «ambiente/macchina» che si basa sull'abbattimento delle emissioni, sulla sostenibilità economica e sociale, sulla gestione intelligente di una macchina o di una flotta di Skid o miniescavatori.) (polocento)
- L'efficace gestione dei rapporti tra sistema delle imprese e della ricerca costituisce pertanto la leva strategica principale per il raggiungimento degli obiettivi. (nanoxm, mentre descrive il "come lavoriamo")
- Gestione del processo di innovazione applicato sul prodotto, sulle tecnologie e su modelli di business (otir2020)
- Gestione dei rapporti esterni con centri e laboratori accreditati, veicolazione di conoscenza (otir2020 entrambe nella sezione "I servizi offerti da OTIR2020 sono:")

GESTORE

Il termine *gestore* è stato scelto perché rilevante alla tematica, ma i dati dimostrano che è usato con bassissima frequenza negli 11 siti e che viene esclusivamente utilizzato in varianti della collocazione *soggetto gestore* per rifersi al gestore del polo che scrive.

IMPRESA

Solamente 3 poli (innopaper, polopenta, polopierre) utilizzano *impresa* all'interno di testi prodotti da loro stessi.Bisogna comunque notare come anche questi 3 poli riportino il termine principalmente all'interno di testi tratti da siti terzi, mentre l'utilizzo all'interno di siti propri è poco frequente – di seguito alcuni dei pochi esempi presenti nei corpora. In tutti i casi è usato all'interno di testi che promuovono l'attività del polo, sottolineando i servizi sviluppati per "l'impresa".

- Il frutto delle attività di auditing tecnologico può concretizzarsi nella progettazione, in stretta collaborazione con l'impresa, di attività di R&S&I (innopaper)
- rafforzamento relazioni Ricerca-Impresa e incubatori d'impresa; (innopaper, parlando del loro progetto Forepaper)
- La consulenza è una modalità trasversale con cui PENTA risponde all'esigenza di miglioramento competitivo della singola impresa, avvalendosi di competenze di eccellenza. (polopenta)
- L'adesione è gratutita, non comporta alcun obbligo da parte di un'impresa (polopierre)

IMPRESE

Imprese è utilizzato da 6 poli (innopaper, distrettoict-robotica, nanoxm, otir2020, polo12, polopenta) in testi il cui scopo è principalmente quello di promuovere i servizi rivolti alle imprese, sia descrivendo ciò che il polo "può fare", sia descrivendo cio che il polo "ha fatto" per migliorare/supportare "le imprese". Di seguito la quasi totalità degli esempi presenti nei corpora dei 6 siti; nonostante il suo utilizzo molto "diretto", *imprese* appare principalmente all'interno di testi tratti da siti terzi, e pertanto le occorrenze dell'uso più "diretto" sono poche.

- ciascuna azienda può accedere ai Servizi offerti dal Polo e finalizzati a promuovere l'accesso delle imprese a servizi specialistici (innopaper)
- Benchmarking dei processi aziendali, scouting e auditing nelle imprese della filiera cartaria (innopaper)
- continua attività di auditing tecnologico al fine di conoscere le caratteristiche delle imprese e di individuare e valutare (innopaper)
- Il nostro obiettivo, far diventare le imprese maggiorenni (distrettoict-robotica; UNA DELLE POCHE occorrenze di uso "diretto", è il titolo di una news che non ha corpo!)
- Buone idee per le imprese del futuro (distrettoict-robotica; VEDI SOPRA)
- con la volontà di facilitare sia la collaborazione tra le imprese stesse, che un rapporto continuativo con i produttori di know how della nostra regione. (distrettoict-robotica; usato nella presentazione del polo)
- composto da 3 centri servizi per le imprese e 4 centri di ricerca, che ne rappresentano oggi il Soggetto Gestore, ed alla adesione di 50 imprese toscane (nanoxm)
- Il Polo Regionale di Innovazione per le Nanotecnologie (NANOXM) mette a disposizione delle proprie imprese aderenti (nanoxm)
- Accompagnare le imprese verso servizi specialistici ad alto valore aggiunto (otir2020)
- materiali di approfondimento tecnico prodotti dai partner di OTIR2020 e messi a disposizione delle imprese per favorirne l'accrescimento di competenze (otir2020)
- il risultato di attività di Ricerca e di Innovazione immediatamente misurabili e utilizzabili da parte delle imprese coinvolte. (otir2020)
- Tutte le imprese toscane appartenenti alla filiera Moda possono aderire gratuitamente e senza alcun impegno al Polo OTIR2020 (otir2020)
- Le imprese aderenti a P12 possono ricevere un contributo a fondo perduto regionale una tantum all'80%. (polo12)
- AL fine di poter identificare le esigenze delle imperese, le linee di sviluppo, gli interessi il personale di P12 visiterà le imprese aggregate (fase di scouting). Questa fase è di estrema

importanza per poter personalizzare l'offerta del polo alle esigenze delle singole imprese (polo12)
- *accompagnare le imprese all'accesso di servizi specialistici ad alto valore aggiunto per sostenere la diffusione dell'innovazione fra le imprese del Polo e le imprese esterne; * facilitare l'accesso da parte delle imprese alla conoscenza scientifica e tecnologica (polopenta)
- PENTA fornirà alle imprese della nautica le proprie competenze ed i propri supporti operativi (polopenta)
- secondo un sistema produttivo fondato su l'eccellenza di un know-how che vede la presenza, al tempo stesso, di grandi cantieri e di numerosissime piccole imprese operanti nella produzione di componenti e servizi specializzati per l'allestimento e la gestione di un mega yacht. (polopenta)
- Con questo servizio PENTA fornisce supporti alla gestione aziendale per migliorare il livello qualitativo delle imprese di filiera. (polopenta)
- Migliorare la visibilità delle imprese e promuovere il miglioramento delle loro relazioni

INDUSTRIALE

Il termine *industriale* è principalmente usato all'interno di testi non prodotti dai poli analizzati; solo 3 poli (innopaper, optoscana, polocento) presentano alcune occorrenze del termine utilizzato in testi relativi alla presentazione dei servizi offerti dai poli stessi. In tutti i casi *industriale* è comunque parte di costruzioni fisse quali *sviluppo industriale, ottica industriale, ricerca industriale, sicurezza industriale*.

- *Strategie di sviluppo industriale (innopaper; nome di una sezione del sito)
- OPTOSCANA si prefigge di aggregare ed organizzare l'attuale eccellenza tecnologica Toscana nei settori dell'Optoelettronica, della Fotonica, dell'Ottica Industriale (optoscana)
- Ambito Tecnologico: scouting, studi di fattibilità, prototipazione, ricerca industriale e precompetitiva, sperimentazioni, validazioni, certificazioni. (optoscana)
- Sicurezza industriale e protezione ambientale. Sicurezza del trasporto di materiali pericolosi. (polocento; descrizione dei servizi offerti)

INNOVAZIONE

Innovazione appare all'interno dei corpora principalmente nelle costruzioni *polo d'innovazione* e *poli di innovazione*, utilizzati dai poli per "descrivere"/riferirsi a se stessi o a poli terzi. Molto frequenti sono anche le collocazioni *ricerca e innovazione, Premio "Impresa + Innovazione + Lavoro"*. È interessante notare come un termine così centrale al concetto stesso di "polo" venga utilizzato in maniera "diretta" e con scopo diverso da quello semplicemente descrittivo (come negli esempi riportati sopra) solo da 3 poli, e sempre con frequenze basse – gli altri poli riportano testi presi da siti terzi, riferiti principalmente a bandi e a premi in cui *innovazione* appare in maniera descrittiva nelle collocazioni riportate in precedenza. Gli esempi riportati qui sotto rappresentano le uniche occorrenze di *innovazione* utilizzato per promuovere la/le attività dei poli. Nel caso delle sezioni del sito, la dicitura appare in ogni pagina del sito.

- Molte delle imprese aderenti ad INNOPAPER hanno potuto godere dei servizi erogati dal Polo e riguardanti attività di innovazione di prodotto e di processo, (innopaper)
- * Innovazione di processo* Innovazione di prodotto (innopaper; nomi di due sezioni del sito)
- * Servizi per l'innovazione (distrettoict-robotica; nome di una sezione del sito)
- una rete di soggetti, attività e servizi per Fare emergere la DOMANDA di INNOVAZIONE proveniente dalle imprese (distrettoict-robotica; nella sezione "chi siamo, cosa facciamo". 1 sola occorrenza)
- Vuoi misurare il tuo grado di innovazione? Gli esperti del Polo sono disponibili ad incontrarti (otir2020)

POLI

Quasi esclusivamente usato in maniera descrittiva negli 11 siti per indicare eventi e/o progetti, e con bassa frequenza (tranne che in polis-toscana.it). È quasi sempre utilizzato nella collocation *poli d'innovazione*, che non è però presente in scienzedellavita.it dove viene usato in *poli e incubatori*. Alcuni siti (innopaper.lucense.it, polopenta.it) descrivono

cosa significhi *poli d'innovazione*. Solo nanoxm.it usa il termine per riferirsi ad attività da esso svolte in concomitanza con optoscana.net "Contributo dei Poli NANOXM e OPTOSCANA al documento regionale".

POLO

Polo è quasi esclusivamente utilizzato come riferimento al polo proprietario del sito, nelle costruzioni *polo d'innovazione NOME*. Appare anche frequentemente in nomi propri quali "Consorzio Polo Tecnologico Magona" (polopenta.it). È prevalentemente usato per descrivere attività/eventi del polo stesso, e raramente utilizzato (polopierre.it, scienzedellavita.it) per riferirsi a poli terzi. Solo polocento.it utilizza *polo* per riferirsi ad un *polo universitario*.

PROCESSI

Il termine, quando non utilizzato all'interno di testi (bandi, notizie, comunicati) riportati da siti terzi, è utilizzato per riferirsi a *processi* “fisici”, “meccanici” - ovviamente in linea con la natura dei poli. È principalmente legato all'idea di “miglioramento”, dove i servizi offerti dal polo permettono di “migliorare” e “supportare” i diversi *processi*. Di seguito è riportata la quasì totalità degli utilizzi del termine all'interno dei siti che ne fanno un uso “diretto”, che rimane comunque sempre poco frequente.

- Industria e innovazione: potenziamento del Laboratorio Centro Qualità Carta per ricerca e sperimentazione di nuovi prodotti; efficienza dei processi (benchmarking, manutenzione, magazzini, logistica, etc) (innopaper; descrizione del loro progetto FOREPAPER)
- Il Polo di Innovazione INNOPAPER lavora per rendere disponibili alle aziende aderenti sia le competenze che l'assistenza tecnologica a supporto dei processi d'innovazione (innopaper; presentazione del polo)
- Competenze e assistenza tecnologica a supporto dei processi d'innovazione che mirano ad incrementare la competitività delle imprese delle filiera cartaria. (innopaper; presentazione del polo)
- Altri dispositivi optoelettronici rivestono ruoli cruciali nei processi di lavorazione industriale di precisione, come i laser di potenza, o nella sensoristica di impiego biomedicale ed ambientale. (optoscana; descrizione della tecnologia OPTOELETTRONICA)
- ricerca relativa alla psicologia dei processi cognitivi ed in particolare ai meccanismi di apprendimento, percezione, attenzione, memoria, ragionamento, e sviluppo cognitivo. In particolare processi decisionali e percezione del rischio, aspetti affettivi e aspetti cognitivi dei processi decisionali, sviluppo e processi decisionali, formazione e apprendimento, comunicazione e uso di tecnologie in ambito sanitario, analisi e definizione di contesti formativi per l'apprendimento. (polocento; descrizione dei servizi offerti)
- La capacità di lavorare in rete, mettendo a sistema conoscenze, competenze e servizi, in particolare in sistemi imprenditoriali e settorialicaratterizzati da piccole dimensioni, è una delle principali opportunità di sviluppo del territorio, poiché agevola l'accesso all'innovazione, stimola i processi collaborativi e rafforza le capacità progettuali e commerciali. (distrettoictrobotica; presentazione del polo)
- Il laboratorio si estende su una superficie complessiva di circa 60 mq ed è diviso in due aree, una per la ricerca e la sperimentazione dei processi ad umido ed una per la ricerca e la sperimentazione dei processi a secco. (nanoxm; descrizione dei servizi offerti)
- Studio di catalizzatori e relativi processi per l›utilizzo energetico di rifiuti e residui agroalimentari. (nanoxm; descrizione dei servizi offerti)
- impartire resistenza agli agenti chimici ed alla radiazione ultravioletta mediante processi a basso impatto ambientale. (nanoxm; descrizione dei servizi offerti)
- facile parallelizzazione e velocizzazione dei processi, abbattimento dei costi con l'utilizzo di dispositivi usa e getta fabbricati con produzione di massa. (nanoxm; descrizione dei servizi offerti)
- In questa sezione è possibile scaricare i materiali di approfondimento tecnico prodotti dai partner di OTIR2020 e messi a disposizione delle imprese per favorirne l'accrescimento di competenze, e la disponibilità ad intraprendere processi di innovazione. Affrontare un processo di ricerca sul prodotto, o innovare i processi produttivi e organizzativi, richiedono in-

fatti una complessa articolazione di conoscenze tecniche e scientifiche che non è possibile mettere in campo autonomamente. (otir2020; presentazione del polo e dei servizi)
- Con questo tipo di servizio PENTA intende fornire una conoscenza aggiornata del settore in termini di processi della produzione, specializzazioni di impresa e servizi al diportismo come supporto per servizi finalizzati (polopenta; presentazione servizi offerti)

PROGETTI

Il termine progetti appare quasi esclusivamente nei testi di presentazione di 6 poli, in cui vengono descritti i servizi offerti e i progetti a cui il polo ha lavorato o la tipologia di progetti nella quale il polo è specializzato. Solo nel sito di polocento il termine appare come titolo di due report relativi – appunto – ai progetti realizzati. Nei rimanenti siti il termine appare sempre all'interno di testi scritti da terzi.

- Inoltre, grazie alle attività svolte ed al supporto fornito alle imprese per lo sviluppo dei loro processi di innovazione, INNOPAPER è stato valutato positivamente ed inserito nell'elenco dei Progetti di Eccellenza in Toscana. I progetti finanziati dal programma PORCReO selezionati come "eccellenze toscane 2013" sono inoltre presenti su youtube (innopaper; descrizione del polo)
- Uno degli strumenti operativi più efficaci a disposizione del Polo OPTOSCANA per favorire le PMI toscane ad esso associate è l'impiego dei Finanziamenti Regionali previsti dal Catalogo dei Servizi Avanzati e Qualificati, in grado di fornire soluzioni tecnologiche per l'innovazione tramite l'attivazione di progetti di ricerca e sviluppo, (optoscana; presentazione del polo)
- * Undici progetti presentati alla scadenza del bando regionale "Ricerca e Sviluppo" (polocento; nome di un report scaricabile)
- Conclusi i progetti di Ricerca e Sviluppo 2012 (polocento; nome di un report scaricabile)
- Tale know-how è stato acquisito sia nell'ambito di progetti di ricerca e sviluppo svolti da singole aziende, (nanoxm; presentazione del polo)
- Apparecchiature tecnico-scientifiche distintive da poter utilizzare per progetti congiunti e/o per consulenze ad imprese (nanoxm; presentazione polo)
- Il partenariato di OTIR2020 è costituito da centri servizi e organismi di ricerca che condividono l'obiettivo comune di collaborare a progetti di innovazione a supporto del sistema produttivo toscano. (otir2020; presentazione del polo e dei suoi partner)
- SERVIZI DI SUPPORTO per la partecipazione a bandi, progetti e iniziative che nascono a livello regionale, nazionale, europeo e internazionale. (otir2020; presentazione dei servizi)
- Progetti di ricerca industriale e/o sviluppo sperimentale (otir2020; presentazione dei servizi)
- SVILUPPO DI PROGETTI e EROGAZIONE DI SERVIZI, P12 collabora con le imprese nella realizzazione di progetti di ricerca e innovazione. (polo12; presentazione polo e servizi)
- Il Polo vive se gli aggregati se è in grado di svolgere attività nel territorio, sviluppando progetti, trasferendo competenze ed erogando servizi qualificati. (polo12; presentazione del polo, sezione "Cosa chiediamo agli aggregati")

PROGETTO

Il termine *progetto* è utilizzato da 4 poli per riferirsi in maniera diretta ad uno specifico progetto sviluppato dal polo. Interessante notare come il polo otir2020 stesso venga indicato come un *progetto* nel sito di otir2020.

- LUCENSE, in collaborazione con il Consorzio INSTM ha avviato il progetto PulPlast, finalizzato al raggiungimento dei seguenti tre obiettivi generali (innopaper)
- Il progetto è promosso e sostenuto proprio dal Distretto Cartario di Capannori, che ha confermato in questo la volontà di valorizzare le proprie radici culturali. (innopaper)
- L'obiettivo generale del progetto è il miglioramento della competitività del sistema di trasformazione della carta tissue (innopaper)
- Il progetto Green Home è finalizzato alla definizione di un Marchio di Qualità ambientale (polocento)
- Il progetto mira a valorizzare le competenze di ricerca industriali e le eccellenze delle regioni italiane e a stimolare le imprese a innovarsi e crescere. (distrettoict-robotica)
- Oltre ai partner dell'Associazione Temporanea di Scopo (ATS), creata per formalizzare il soggetto gestore del progetto OTIR2020 (otir2020)

- Attività di sperimentazione gratuite su prodotti di interesse presso i laboratori che vengono messi a disposizione dei partner del progetto OTIR2020. (otir2020)

RICERCA

Tra i termini analizzati, *ricerca* è quello che appare con frequenza maggiore in tutti gli 11 siti. È probabilmente anche il termine più interessante per quanto riguarda la costruzione del significato: *ricerca* può infatti riferirsi a concetti molto diversi tra loro, che possono essere riassunti in:

- significato di "cercare qualcosa/qualcuno", come in *ricerca di aziende* (polopenta.it)
- significato di "analisi di dati", come in *ricerca di settore* (otir2020.it)
- significato di "ricerca accademica", come in *finanziamento alla ricerca europea, incentivi alla ricerca, trainata dalla ricerca* (distrettoict-robotica.it)
- significato descrittivo di un settore d'industria, *ricerca e sviluppo*

È stato escluso il significato di "cercare all'interno del sito" (il pulsante *ricerca*), in quanto inutile ai fini dell'analisi linguistica, e quasi completamente eliminato dai dati durante la pulizia dei corpora.

È interessante notare come i quattro significati sopraindicati siano ben definiti, e se in un sito *ricerca* appare con il terzo significato, difficilmente appare con il primo e/o il secondo. Il primo e il secondo significato sono frequentemente presenti insieme. Fa eccezione il quarto significato, in quanto è parte di una collocation specifica e molto frequente.

Si può pertanto notare come siano diversi gli atteggiamenti nei confronti della ricerca – e cosa si intenda con essa – negli 11 siti:

otir2020.it, scienzedellavita.it, polis-toscana.it, polopenta.it presentano prevalentemente un utilizzo del primo e del secondo significato, con un'altissima frequenza della collocation *ricerca e sviluppo* all'interno di titoli o diciture all'interno di contenuti riportati da siti terzi.

innopaper.lucense.it, optoscana.net, distrettoict-robotica.it, nanoxm.it, polo12.it, polopierre.it vedono la *ricerca* principalmente come un'attività accademica, di importanza per la creazione e il miglioramento delle attività. È interessante notare come sia quasi del tutto assente la collocation *ricerca e sviluppo*.

polocento.it utilizza quasi esclusivamente la collocation *ricerca e sviluppo*, e un'alta frequenza del termine nel nome proprio *Centro ricerca ed energia* e *Servizi di Ateneo per la Valorizzazione della Ricerca e la Gestione dell'Incubatore Universitario.* A quest'ultimo sono collegate le occorrenze di *progetti di ricerca* e *laboratori di ricerca*.

SERVIZI

Usato all'interno di testi scritti dal polo proprietario del sito (in 5 siti), il termine *servizi* appare (non inaspettatamente) per lo più nella sezione in cui il polo presenta i propri *servizi*. Interessante notare come il numero di siti in cui i servizi vengono presentati sia ridotto.

- Attraverso l'adesione al Polo di Innovazione INNOPAPER, ciascuna azienda può accedere ai Servizi offerti dal Polo e finalizzati (innopaper; presentazione polo/servizi)
- Molte delle imprese aderenti ad INNOPAPER hanno potuto godere dei servizi erogati dal Polo (innopaper)
- Per questa ultima tipologia di servizi, INNOPAPER si avvale delle attrezzature tecnico-scientifiche (innopaper)
- i Servizi Qualificati erogati dal Polo Cento I risultati, a novembre 2013, del lavoro svolto sul bando regionale “Aiuti alle piccole e medie imprese dell'industria, dell'artigianato e dei servizi alla produzione per l'acquisizione di servizi avanzati e qualificati” (polocento; presentazione servizi)

- Tra i Servizi Qualificati erogati dal Polo Cento ci sono i “Servizi di supporto alla innovazione di prodotto nella fase di concetto”, con sei progetti seguiti, i “Servizi di supporto all'introduzione di nuovi prodotti”, con quattro progetti seguiti, i “Servizi tecnici di progettazione per innovazione di prodotto e di processo produttivo”, (polocento)
- Il Polo ICT e robotica rappresenta un punto di aggregazione e garantisce l'accesso a servizi, competenze e infrastrutture tecnologiche con una copertura di servizi all'intero (distrettoict-robotica; presentazione polo/servizi)
- Il Polo di Innovazione ICT-Robotica si propone come partner qualificato delle imprese hi-tech per l'erogazione di servizi qualificati tra loro integrati e complementari. (distrettoict-robotica)
- Competenze e servizi know-how specifico relativo ad applicazioni trasversali di nanotecnologie (nanoxm; presentazione polo, c'è un'intera sezione del sito chiamata “Servizi”)
- Un contributo sino all'80% sull'acquisizione di servizi di Consulenza Qualificata per le pmi toscane dell'industria (otir2020)
- Servizi e opportunità I servizi offerti da OTIR2020 sono: SERVIZI DI NATURA TECNOLOGICA: (otir2020; presentazione servizi)
- PENTA fornirà alle imprese della nautica le proprie competenze ed i propri supporti operativi per favorire il loro accesso ai seguenti servizi revisti da questa linea di finanziamento: * Gestione della catena di fornitura o supply chain * Servizi per l'efficienza ambientale ed energetica (polopenta)

SERVIZIO

È interessante notare come il termine *servizio*, molto meno frequente del plurale *servizi*, sia usato raramente con il significato del plurale *servizi*. Il *servizio* non è pertanto il singolare delle “attività, esperienze” offerte dal polo, ma è più che altro utilizzato all'interno della costruzione “al servizio di”. Solo polocento utilizza il termine *servizio* con il significato più concreto del plurale *servizi*.

- Il Polo di Innovazione INNOPAPER, anche grazie all'esperienza del soggetto gestore LUCENSE che da anni opera a servizio delle imprese (innopaper; presentazione polo, solo 3 occorrenze totali)
- INNOPAPER che da oltre 30 anni opera a servizio delle aziende della filiera cartaria
- A questi si aggiungono i “Servizi di supporto al cambiamento organizzativo”, con quattro servizi erogati, i “Servizi di miglioramento della efficienza delle operazioni produttive”, con due servizi erogati, i “Servizi di supporto all'innovazione organizzativa mediante gestione temporanea di impresa”, con un servizio erogato (polocento; presentazione servizi)
- Il Soggetto Gestore di NANOXM è oggi rappresentato da una ATS di organismi di ricerca e centri di servizio aderenti alla TECNORETE (nanoxm; presentazione polo)
- È a questo tipo di imprese che si rivolge P12 offrendo servizi basati su una struttura associativa con competenze differenziate (Centri di Servizio, Enti di Ricerca, (polo12; presentazione polo)
- Il servizio in particolare è finalizzato a fornire servizi per una gestione innovativa delle commesse da parte di singole imprese o loro specifici raggruppamenti per: (polopenta; presentazione polo)
- Con questo servizio PENTA fornisce supporti alla gestione aziendale (polopenta; presentazione polo)
- Il Polo di Innovazione PIERRE offre ai propri aderenti un servizio di supporto (polopierre; presentazione polo)

SVILUPPO

Come visto in precedenza, il termine è strettamente legato a *ricerca* nella collocation *ricerca e sviluppo*, che è maggiormente frequente in otir2020.it, scienzedellavita.it e polistoscana.it.

Otir2020.it, scienzedellavita.it e polocento.it utilizzano prevalentemente il termine per riferirsi a oggetti materiali quali *sviluppo di nuovi alimenti funzionali* (scienzedellavita.it), *sviluppo del prodotto* (otir2020.it), *sviluppo di prototipi/codici/sette progetti* (polocento.it)

innopaper.lucense.it, distrettoict-robotica.it, nanoxm.it, polo12.it, polopenta.it e polopierre.it intendono invece lo *sviluppo* come riferito a qualcosa di più immateriale/processuale: *strategia di sviluppo* (innopaper.lucense.it), *servizi per accelerare lo sviluppo delle aziende toscane* (distrettoict-robotica.it), *motore dello sviluppo* (nanoxm.it), *sviluppo e competitività, non-sviluppo* (polo12.it), *importanti scommesse sul proprio sviluppo futuro* (polopenta.it), *ottimizzazione dello sviluppo* (polopierre.it).

optoscana.net è invece l'unico sito che utilizza il termine nel nome proprio *Sviluppo Toscana* e con una prevalenza della collocation *ricerca e sviluppo*.

Polis-toscana.it riporta invece il termine all'interno di titoli di contenuti tratti da siti terzi.

TECNOLOGIA

Il termine è principalmente usato negli 11 siti come parte di costruzioni quali *Scienza e tecnologia dei materiali*, oppure *tecnologia rfid/laser/NOME* (ad indicare una specifica tecnologia legata ad un prodotto). È usata in testi scritti dal polo proprietario del sito in 3 siti, e solo otir2020 non lo utilizza per riferirsi ad una tecnologia precisa:

- Conclusa la fase di studio di fattibilità dell'impianto di trattamento dello scarto di pulper (lo scarto derivante dai processi di riciclaggio della carta) con tecnologia torcia al plasma, LUCENSE ha avviato (innopaper)
- Infatti la soluzione “recupero di materia”, non implicando il ripudio della tecnologia finalizzata al “recupero di energia” (innopaper)
- rispetto ai metodi tradizionali, questa tecnologia offre molti vantaggi (nanoxm)
- Gli esperti del Polo sono disponibili ad incontrarti presso la vostra azienda per discutere delle problematiche del sistema aziendale e metterti a conoscenza delle più recenti soluzioni in termini di tecnologia ed innovazione di interesse per la tua attività (otir2020; presentazione servizi)

TECNOLOGIE

Tecnologie è più frequente nei corpora del singolare *tecnologia*, principalmente perché utilizzato nel nome del polo (ad es. Polo dell'innovazione della nautica e delle tecnologie del mare) e come parte di costruzioni quali *tecnologie meccaniche/industriali/optoelettroniche*. Viene utilizzato all'interno della sezione di presentazione del polo e/o dei servizi da 3 poli.

- Tecnologie di valorizzazione dei rifiuti cartari (innopaper\presentazione dei servizi)
- contribuendo così in maniera effettiva al trasferimento di tecnologie, alla messa in rete e alla diffusione delle informazioni tra le imprese che costituiscono il Polo stesso (innopaper\presentazione del polo)
- sono potenzialmente applicabili in campi analitici per tecnologie che sfruttano proprietà magnetiche, ottiche od entrambe, come avviene nel campo delle applicazioni biochimiche e biomediche. (nanoxm\presentazione servizi)
- Tecnologie a scansione di sonda (nanoxm\presentazione servizi)
- nonché contribuendo in maniera effettiva al trasferimento di tecnologie, alla messa in rete e alla diffusione delle informazioni tra le imprese che costituiscono il Polo (nanoxm\presentazione polo)
- Strumenti e tecnologie per il design innovativo, la riduzione dell'impatto ambientale, l' incremento dell'efficienza e il risparmio di risorse energetiche e ambientali (otir2020\presentazione servizi)
- Gestione del processo di innovazione applicato sul prodotto, sulle tecnologie e su modelli di business (otir2020\presentazione servizi)

TERRITORIO

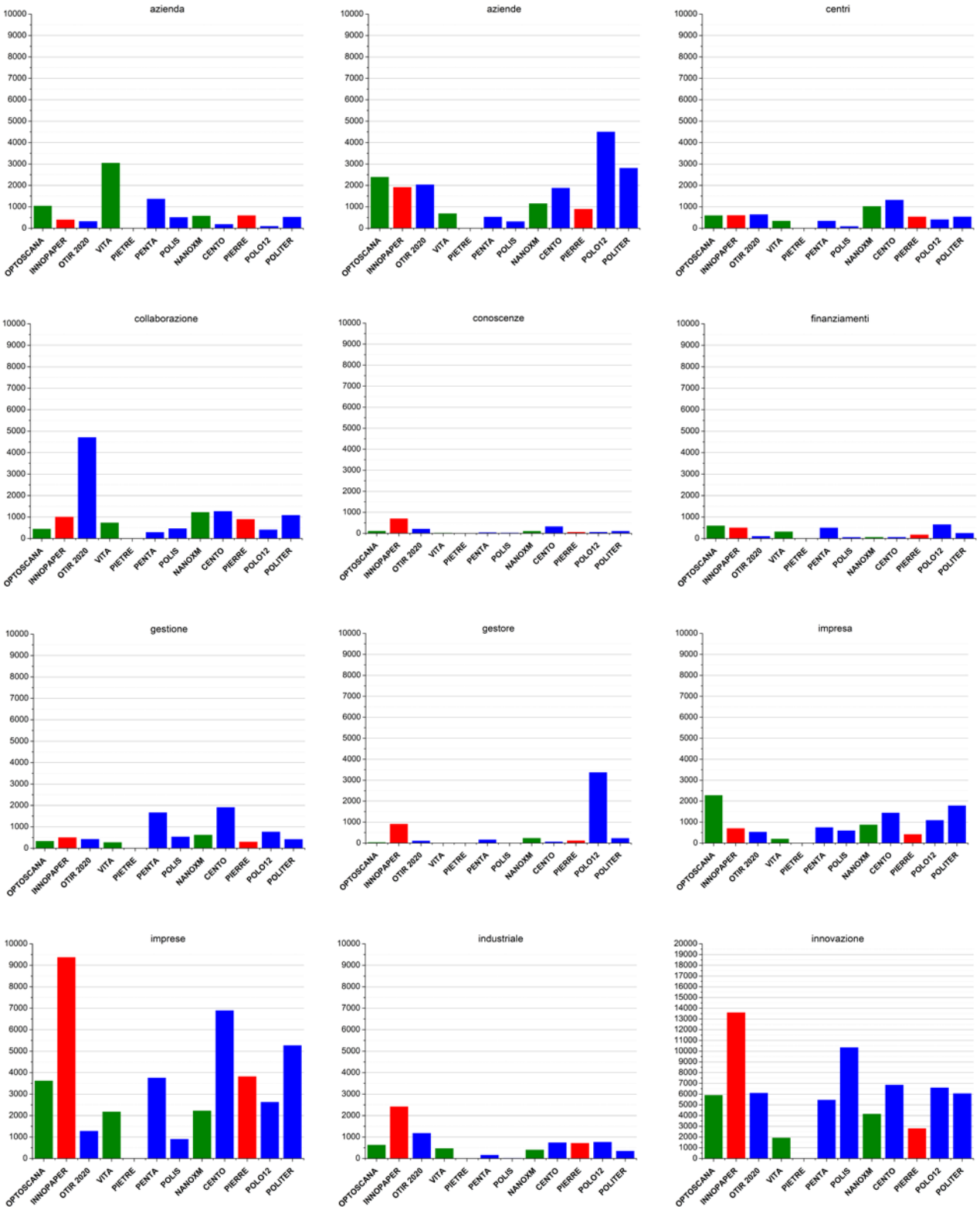
Il termine *territorio* appare con basse frequenze in tutti gli 11 siti; ciononostante è interessante notare come vi siano due distinte costruzioni del termine. Per la maggior parte dei siti il termine è sempre relativo allo spazio circostante il polo, o allo spazio di cui il polo fa parte, ed appare pertanto in occorrenze quali *aziende del territorio, attrattività del*

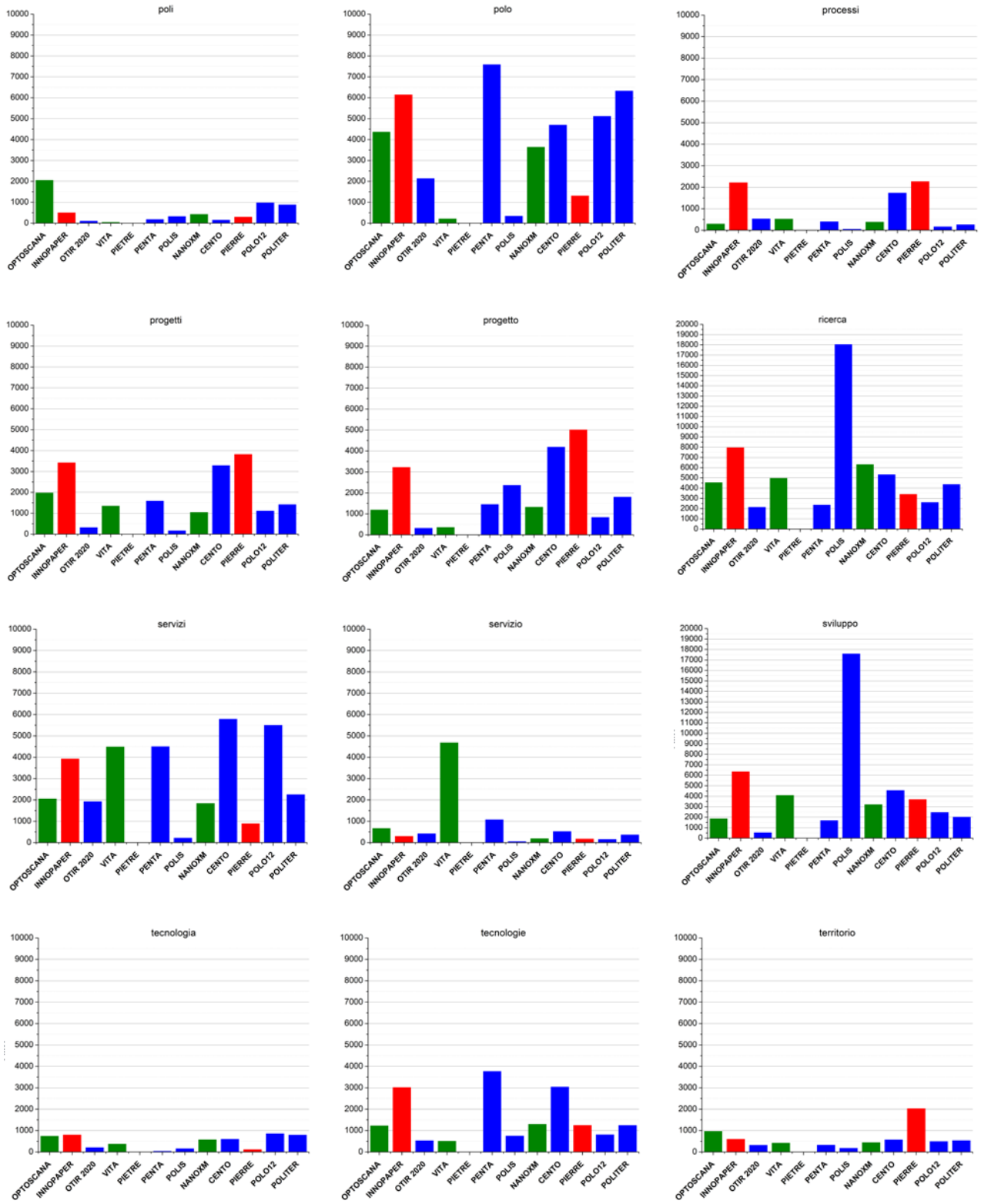
territorio (nanoxm.it), *Le politiche dell'Innovazione del territorio nell'area metropolitana* (distrettoict-robotica.it), *Le grandi imprese presenti sul territorio* (optoscana.net), *radicamento nel territorio* (polocento.it)

In generale il concetto di *territorio* è costruito in relazione ad un'idea più ampia, non legata solo allo spazio, e in funzione di progetti per la salvaguardia/protezione dello stesso.

Figura 2 Occorrenze dei 24 termini selezionati dai siti web analizzati

frequenze normalizzate per milione di parole.
Legenda colori istogramma: poli di fascia 1: blu; fascia 2: rosso; fascia 3: verde





Fonte: ns. elaborazione su dati tratti dai siti web dei poli di innovazione; estrazione maggio 2015

3 Collegamenti web

Nell'analisi dei testi pubblicati dai siti web dei poli di innovazione abbiamo estratto i link ai domini principali dei siti web dei poli e dei soggetti gestori citati nei siti web dei poli[6]. Gli obiettivi dell'analisi sono due. Un primo obiettivo è capire in che misura attraverso i siti web i poli hanno comunicato le connessioni tra di loro e con i soggetti attivi nel trasferimento tecnologico. Un approfondimento riguarda anche le organizzazioni che hanno stretto accordi di collaborazione con i poli, le società che partecipano nei soggetti gestori, i laboratori e gli incubatori. Un secondo obiettivo è capire in che misura i poli fanno riferimento allo stesso spazio di informazioni. A questo scopo sono stati utilizzati i link completi presenti nei siti web dei singoli poli e si è cercato di identificare i riferimenti comuni. Risponderemo a queste domande nelle sezioni (a) e (b) di questo paragrafo.

Dall'analisi è escluso il polo Pietre che al momento dell'elaborazione (maggio 2015) non era attivo.

a. La comunicazione web come misura delle connessioni tra i poli

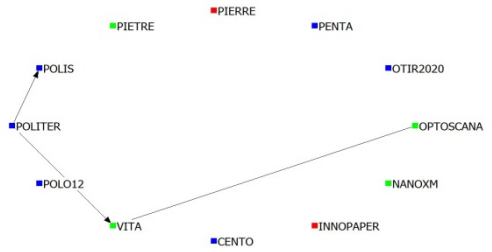
Nella Tabella 2 è riportato il numero di citazioni di siti web di un polo fatte da un altro polo (sulle proprie pagine web). Come si può vedere anche dal Grafo 1[7], sono solo due i poli che citano i siti web di altri poli: Optoscana, che cita Vita, e Politer che cita sia Polis che il sito della Fondazione Toscana Life Sciences (su cui si trova la pagina web del polo Vita).

Tabella 2 Numero di siti web dei poli citati da altri poli

	OPTOSCANA	INNOPAPER	OTIR2020	VITA	PIETRE	PENTA	POLIS	NANOXM	CENTO	PIERRE	POLO12	POLITER
OPTOSCANA				1								
INNOPAPER												
OTIR2020												
VITA												
PIETRE												
PENTA												
POLIS												
NANOXM												
CENTO												
PIERRE												
POLO12												
POLITER				1			1					

Grafo 1 Reti di relazioni tra i poli attraverso la citazione dei siti web di altri poli

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3. Sono omessi i loop



Nella Tabella 3 è riportato il numero di volte in cui nel sito web di un polo viene citato un soggetto gestore, che può essere del polo stesso o di un altro polo. Ad esempio, il polo Optoscana ha citato nel proprio sito web il link di un soggetto gestore di Otir2020. Il polo Innopaper ha inserito nel suo sito web il link al suo (unico) soggetto gestore e ha citato anche un gestore di Polis, uno di Nanoxm, uno di Pierre e uno di Politer. Solo il polo Cento non ha indicato sul proprio sito dei link alle pagine web di soggetti gestori .

[6] Non è stato possibile raccogliere informazioni sul polo Pietre: nonostante ufficialmente esista un sito dedicato, non è possibile accedervi. Pertanto, manca completamente da questa analisi (e, ovviamente, il polo Pietre non risulta avere nessuna relazione con gli altri poli). Inoltre è stata fatta un'operazione di pulizia per eliminare i domini dei siti web inutili ai fini dell'indagine (docs.google, wordpress, youtube, flyby, etc.).

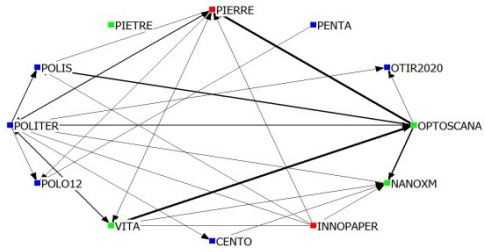
[7] I collegamenti del Grafo 20 hanno una direzionalità, indicata dalla punta della freccia, perché la citazione viene fatta da un polo verso l'altro, ma non è detto che ciò avvenga anche nel senso opposto, come si può intuire dal fatto che la matrice della Tabella 17 non è simmetrica.

Tabella 3 Numero di siti web di soggetti gestori citati da poli

	OPTOSCANA	INNOPAPER	OTIR2020	VITA	PIETRE	PENTA	POLIS	NANOXM	CENTO	PIERRE	POLO12	POLITER
OPTOSCANA	2		1	5			3	3		5		2
INNOPAPER		1		1			1	1		1		1
OTIR2020			6									
VITA	1			1				1		1		1
PIETRE												
PENTA						1					1	
POLIS							1					
NANOXM								3	1			1
CENTO												
PIERRE				1			1			1	1	1
POLO12											6	
POLITER	1		1	2			2		1	2	1	8

Grafo 2 Reti di relazioni tra i poli attraverso la citazione dei siti web dei soggetti gestori

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3. Sono omessi i loop



Eccezion fatta per il polo Pietre (mancante nella nostra analisi), la direzionalità dei collegamenti riportata nel Grafo 2 evidenzia che i poli Otir2020, Polis, Cento e Polo12 non citano soggetti gestori di altri poli (ma alcuni dei loro gestori sono citati da altri poli), mentre il polo Optoscana e il polo Politer sono quelli che citano più volte soggetti gestori di altri poli. In particolare Optoscana cita 5 soggetti gestori del polo Vita (si tratta delle università toscane: UNIFI, UNIPI, UNISI, Scuola Normale e Sant'Anna) e 5 soggetti del polo Pierre (UNIFI, UNIPI, UNISI, Sant'Anna e ASEV).

Nella Tabella 4 è riportato il numero di riferimenti a siti web delle organizzazioni che hanno stretto accordi di collaborazione con i poli. Per fare quest'analisi sono state prese in considerazione tutte le organizzazioni che hanno stretto accordi di collaborazione con i poli e si è individuata la presenza di un link al loro sito web nei siti dei poli: in molti casi si è constatato che quelle organizzazioni non sono state citate nei siti web dei poli con cui avevano stretto accordi di collaborazione, anche se sono state citate nei siti web di altri poli. È possibile ricavare questa informazione osservando la diagonale principale della Tabella 4: solamente Optoscana, Innopaper, Vita e Politer hanno citato organizzazioni con le quali hanno stretto degli accordi di collaborazione, mentre tutti gli altri poli (a parte Penta e Cento, che non hanno citato organizzazioni con le quali sono stati stipulati accordi) hanno citato organizzazioni che hanno stipulato accordi di collaborazione con altri poli[8].

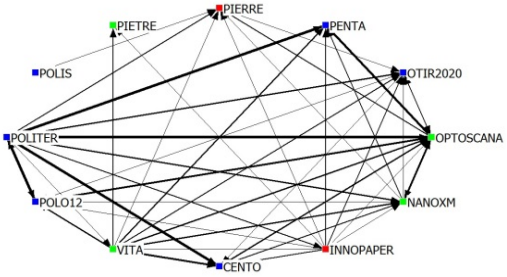
[8] In questa analisi sono considerati tutti gli accordi di collaborazione stipulati.

Tabella 4 Numero di riferimenti a siti web delle organizzazioni che hanno stretto accordi di collaborazione con i poli

	OPTOSCANA	INNOPAPER	OTIR2020	VITA	PIETRE	PENTA	POLIS	NANOXM	CENTO	PIERRE	POLO12	POLITER
OPTOSCANA	11		2	2		8		6	4	2	7	1
INNOPAPER	1	1	2	1	1	2		1	3	1	1	
OTIR2020	1								1		1	
VITA	5		3	8	2	4		4	7	1	4	
PIETRE												
PENTA												
POLIS										1		
NANOXM	2		1	1		2			2	1	1	
CENTO												
PIERRE			1									
POLO12			1									1
POLITER	10	2	4	1		10		5	9	3	10	4

Grafo 3 Reti di relazioni tra i poli attraverso la citazione dei siti web delle organizzazioni che hanno stretto accordi di collaborazione con i poli

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3. Sono omessi i loop



Dal Grafo 3 si può vedere come le relazioni più importanti (legate al numero di link presenti nei siti web dei singoli poli) passino tutte attraverso Politer. Si tratta però di relazioni molto asimmetriche, perché Politer, ad esempio, cita spesso organizzazioni che hanno stretto accordi di collaborazione con Optoscana, con Penta e con Polo12, mentre questi ultimi citano pochissime volte (o addirittura mai) organizzazioni che hanno stretto accordi di collaborazione con Politer.

Nella Tabella 5 è possibile vedere quante volte un polo cita nel suo sito il link alla pagina web dei propri laboratori o di laboratori di altri poli. I primi sono osservabili sulla diagonale della matrice: solo 6 poli su 12 citano almeno una volta i propri laboratori. I secondi invece sono osservabili in tutte le altre celle della matrice; anche in questo caso si tratta di una matrice asimmetrica, per cui è possibile che un polo citi i laboratori di un altro polo ma che non avvenga il contrario.

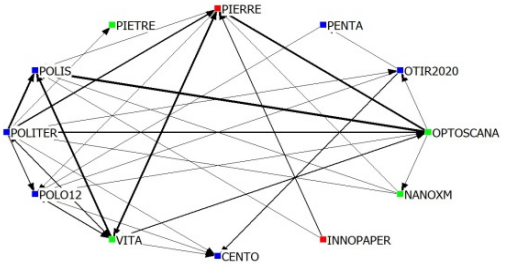
Poiché i laboratori sono un elemento molto importante per i poli di innovazione (un'impresa potrebbe aver deciso se aderire o meno ad un polo in base alla disponibilità di infrastrutture tecnologiche), ci si attendeva che ci fosse un'adeguata informazione on-line. In realtà ciò non è avvenuto: anzi, nel Grafo 4 è possibile cogliere come ci siano stati alcuni casi (Polis su tutti) in cui i laboratori a disposizione di un polo sono stati citati molte più volte da altri poli che da quello a cui facevano riferimento.

Tabella 5 Numero di riferimenti a siti web dei laboratori a disposizione dei poli

	OPTOSCANA	INNOPAPER	OTIR2020	VITA	PIETRE	PENTA	POLIS	NANOXM	CENTO	PIERRE	POLO12	POLITER
OPTOSCANA	5		1	1			7	1		5	1	4
INNOPAPER		1					1			2		
OTIR2020			3			1			2		1	
VITA	3						6		1	6	2	2
PIETRE												
PENTA											1	
POLIS							1					
NANOXM							1			1		1
CENTO												
PIERRE				1			1				1	
POLO12				1					1		4	
POLITER	1		1	2	1		6		1	4	3	5

Grafo 4 Reti di relazioni tra i poli attraverso la citazione dei siti web dei laboratori a disposizione

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3. Sono omessi i loop



Nella Tabella 6 è possibile vedere quante volte un polo cita nel suo sito il link alla pagina web degli incubatori. Solo in tre casi (Otir2020, Polo12 e Politer) vengono citati almeno una volta gli incubatori messi a disposizione dal polo di innovazione (dati sulla diagonale della matrice). Rispetto ai laboratori, il numero di incubatori che appartengono ad uno o

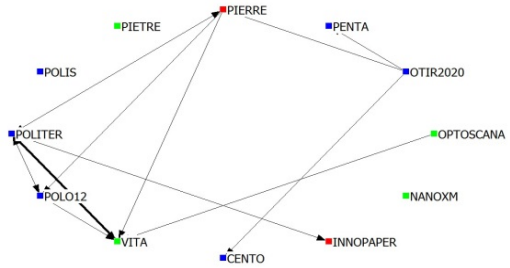
più poli di innovazione è molto ridotto: per questo motivo le citazioni sono molte meno in termini assoluti.

Tabella 6 Numero di riferimenti a siti web degli incubatori

	OPTOSCANA	INNOPAPER	OTIR2020	VITA	PIETRE	PENTA	POLIS	NANOXM	CENTO	PIERRE	POLO12	POLITER
OPTOSCANA				1								
INNOPAPER												
OTIR2020			1			1			1	1		
VITA												
PIETRE												
PENTA												
POLIS												
NANOXM												
CENTO												
PIERRE				1							1	1
POLO12				1							1	1
POLITER		1		2						1	1	2

Grafo 5 Reti di relazioni tra i poli attraverso la citazione dei siti web degli incubatori

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3. Sono omessi i loop



Come si può vedere dal Grafo 5, solo in un caso un polo cita due incubatori di un altro polo: si tratta di Politer, nel cui sito sono presenti i link agli incubatori Toscana Life Sciences (Bio-Incubatore) e Pont-Tech (CERFITT) del polo Vita. I poli che non citano nessun incubatore sono Innopaper, Vita, Penta, Polis, Nanoxm e Cento (Pietre non è considerato nell'analisi). Gli incubatori che afferiscono a questi poli vengono però citati da altri[9].

L'ultima analisi effettuata utilizzando i domini web come fonte di informazione riguarda i partecipanti nei soggetti gestori. In alcuni siti dei poli sono stati individuati link alle pagine web di partecipanti; questi ultimi possono essere presenti in più soggetti gestori, e i soggetti gestori possono essere presenti in più poli. L'obiettivo dell'analisi è individuare il collegamento che unisce un partecipante ad un polo, che può avere valore superiore ad uno perché un partecipante può detenere quote societarie di più gestori nello stesso polo. Una volta costruiti questi collegamenti, è stato possibile collegare tra loro i vari poli, in base alla citazione della pagina web di uno o più partecipanti sui siti dei poli. Come si può vedere nella Tabella 7, i principali poli che riportano link di società partecipanti nei soggetti gestori sono Optoscana, Innopaper, Politer e Pierre. Di questi, Politer e Pierre sono quelli che citano più di una volta società partecipanti nei propri gestori. Anche in questo caso la direzionalità è molto importante: diversi poli citano link a pagine web di società partecipanti nei soggetti gestori del polo Nanoxm, ma quest'ultimo cita solo una volta società partecipanti nei soggetti gestori di altri poli.

Nel Grafo 6 viene visualizzata la rete di relazioni tra i poli attraverso il collegamento “sito web del polo di innovazione – partecipante nel soggetto gestore – soggetto gestore – polo di innovazione”. Le relazioni più intense sono quelle attivate principalmente da Optoscana e Politer.

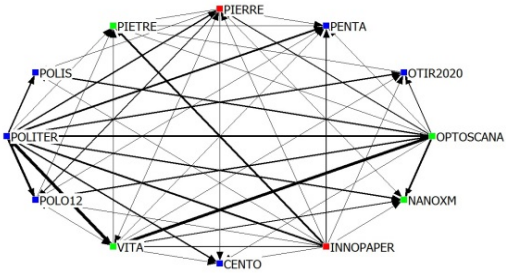
[9] Si tratta dei seguenti casi: l'Incubatore del Polo Tecnologico Lucchese di Innopaper, citato da Politer; la Fondazione Toscana Life Sciences (Bio-Incubatore), Pont-Tech (CERFITT) e Vita, citati da Optoscana, Pierre, Polo12 e Politer; il Consorzio Polo Tecnologico Magona, di Penta, citato da Otir2020 (di cui però fa ugualmente parte, quindi si può dire che in questo caso dipenda dal fatto che questo incubatore sia presente in più poli); il Consorzio Polo Tecnologico Magona di Cento, citato da Otir2020 (di cui però fa ugualmente parte, come detto in precedenza).

Tabella 7 Numero di riferimenti a siti web dei partecipanti nei soggetti gestori dei poli

	OPTOSCANA	INNOPAPER	OTIR2020	VITA	PIETRE	PENTA	POLIS	NANOXM	CENTO	PIERRE	POLO12	POLITER
OPTOSCANA	**1**		2	7	1	1	3	4	1	2	3	4
INNOPAPER		**1**	1	2	4	2	1	1	1	1	1	4
OTIR2020					1							
VITA			1	**1**	1	1		2	1	1	1	1
PIETRE												
PENTA					1							
POLIS												
NANOXM				1				**1**				
CENTO												
PIERRE				1	1	1	1	1	1	**2**	2	
POLO12				1	1	1					**1**	
POLITER	1		3	7	1	4	3	3	3	3	5	**3**

Grafo 6 Reti di relazioni tra i poli attraverso la citazione dei siti web dei partecipanti nei soggetti gestori

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3. Sono omessi i loop



b. La comunicazione web come misura dello spazio di informazione condiviso

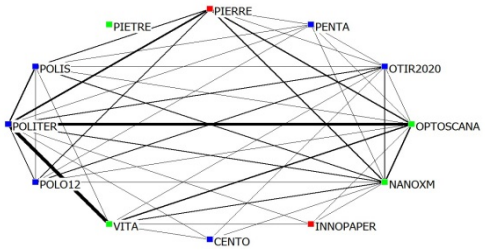
Dall'analisi dei siti web possiamo desumere in che misura i poli facciano riferimento allo stesso spazio di informazioni. Dalla Tabella 8 è possibile vedere quali relazioni intercorrono attraverso i link completi (ovvero specifiche pagine web) presenti nei siti internet. Da tale analisi sono stati esclusi i domini di servizio nella creazione e gestione di pagine web.

Tabella 8 Numero di link completi presenti nei siti internet dei poli (sulla diagonale) e in comune tra i poli

	OPTOSCANA	INNOPAPER	OTIR2020	VITA	PIETRE	PENTA	POLIS	NANOXM	CENTO	PIERRE	POLO12	POLITER
OPTOSCANA	**147**	1	2	10		1	2	10		8	2	24
INNOPAPER	1	**140**		1				1				1
OTIR2020	2		**25**	2		1	2	5	1	1	3	4
VITA	10	1	2	**266**			1	5	1			30
PIETRE												
PENTA	1		1			**87**	1	1		2	2	1
POLIS	2		2	1		1	**50**	3		3	2	9
NANOXM	10	1	5	5		1	3	**100**	1	7	2	5
CENTO			1	1				1	**7**			1
PIERRE	8		1			2	3	7		**204**	3	12
POLO12	2		3			2	2	2		3	**48**	6
POLITER	24	1	4	30		1	9	5	1	12	6	**973**

Grafo 7 Reti di relazioni tra i poli attraverso i link completi presenti nei loro siti web

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3. Lo spessore del legame tra i nodi è proporzionale al numero di link completi ai siti web in comune. Sono omessi i loop



Il numero di link completi condivisi va da un minimo di zero ad un massimo di 30 (tra Vita e Politer). Il polo Innopaper è quello che risulta meno collegato ad altri poli, mentre le relazioni più marcate sono quelle tra Politer, Vita e Optoscana; Pierre risulta avere diverse connessioni con Politer e Optoscana. Anche su questa rete sono state calcolate le misure di centralità (Tabella 9). Nanoxm e Politer hanno i risultati migliori sotto ogni punto di vista.

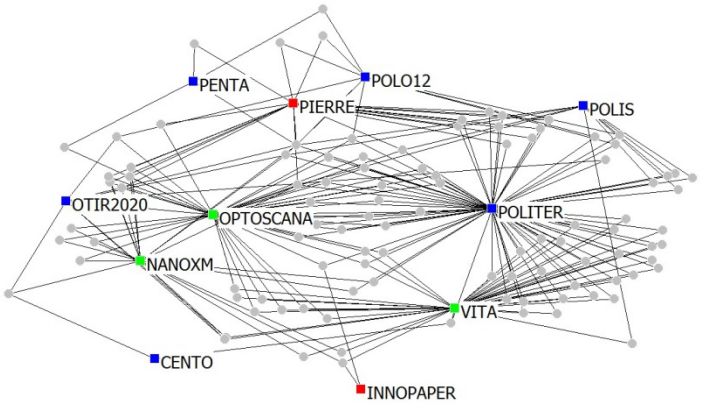
Tabella 9 Misure di centralità dei poli (rete unimodale) sulla condivisione di link completi a pagine web

	Degree	Bonacich Power	Closeness	Eigenvectors	Betweenness
OPTOSCANA	10	1984.84	14	0.35	2.10
INNOPAPER	5	975.67	19	0.17	0.00
OTIR2020	10	1984.84	14	0.35	2.10
VITA	8	1519.96	16	0.26	1.33
PIETRE	0	0.00	33	0.00	0.00
PENTA	8	1703.46	16	0.30	0.00
POLIS	9	1874.62	15	0.33	0.60
NANOXM	11	2095.07	13	0.36	3.93
CENTO	5	975.67	19	0.17	0.00
PIERRE	8	1703.46	16	0.30	0.00
POLO12	8	1703.46	16	0.30	0.00
POLITER	11	2095.07	13	0.36	3.93

Prendendo in considerazione i singoli collegamenti, è possibile creare un grafo bimodale in cui viene presa in considerazione la relazione sito web-polo di innovazione. Il Grafo 8 rappresenta la relazione tra i link completi citati da più poli e la relazione con i poli che li hanno citati. Politer, Vita e Optoscana rivestono un ruolo centrale in questa rappresentazione, ed è interessante notare come i poli di terza fascia (Vita, Optoscana e Nanoxm) siano molto vicini ed interrelati tra loro. Probabilmente ciò dipende dalla specificità dei settori industriali di riferimento, che hanno molti punti in comune, ma è possibile anche che questi poli abbiano usato la rete per fare network e sopperire così alla loro ridotta dimensione.

Grafo 8 Reti di relazioni tra i poli e i link completi presenti nelle pagine web ufficiali dei poli

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3
Lo spessore del legame tra i nodi è proporzionale al numero di soggetti in comune



c. Reti virtuali

La Tabella 10 riporta il numero di relazioni che risulta dall'unione delle reti virtuali. Emerge un quadro di come i poli hanno gestito la comunicazione, sia in riferimento al loro spazio di relazioni formali, che con riferimento allo spazio di attività comuni. I poli più attivi dal punto di vista delle citazioni e delle auto-citazioni sono Optoscana, Innopaper, Vita e Politer, che sono quindi buoni utilizzatori del web nella promozione di tutto ciò che è collegato ai poli di innovazione.

Sulle reti di relazioni virtuali sono state calcolate le misure di centralità per i 12 poli, riportate nella Tabella 11. Gli indici di centralità (Degree, Bonacich Power e Closeness) sono suddivisi in due categorie: “In” e “Out”, per chi riceve e per chi dà una citazione. Per questo tipo di relazioni non è possibile calcolare gli Eigenvectors. L'indice di centralità di Bonacich per chi dà una citazione (Out Bonacich Power) è particolarmente alto per Politer e per Innopaper (così come l'OutDegree), mentre per chi riceve una citazione (In

Bonacich Power) è molto alto per Otir2020 e Cento. Pierre, Polo12 e Politer sono quelli che hanno l'indice di Betweenness più alto.

L'analisi delle cliques sulla rete unimodale delle reti virtuali (Tabella 12 e Tabella 13), mostra un calo di possibili combinazioni rispetto alle reti dei promotori del sistema dei poli e alla rete di competenze. È possibile che la forte polarizzazione, dovuta alla presenza di attori più forti dal punto di vista della comunicazione web, abbia fatto sì che il numero di sottogruppi si riducesse al minimo.

Vi sono diversi elementi interessanti che emergono da questa analisi. Il primo è la presenza costante, in tutte e tre le possibili cliques, dei poli Vita e Politer. Il secondo è l'assenza del polo Innopaper. Questo probabilmente dipende dal fatto che il polo stesso, i suoi gestori, le società partecipanti e le strutture collegate non sono citate da nessun altro polo (a parte Politer): è Innopaper ad essere molto attivo nella citazione di pagine web di attori che appartengono ad altri poli, mentre gli altri poli non citano nulla che riguarda Innopaper. Il terzo è l'assenza del polo Penta: non ha nessuna relazione con attori che appartengono alle prime due cliques, e ha una debole relazione con Polo12, non sufficiente però per poter appartenere alla terza clique (alla quale appartiene Polo12).

Tabella 10 Numero di link a pagine web: dei poli, dei soggetti gestori, delle società con le quali hanno accordi di collaborazione, delle società partecipanti nei soggetti gestori, degli incubatori e dei laboratori dei poli
i link a propri gestori, laboratori, etc. sono riportati sulla diagonale

	OPTOSCANA	INNOPAPER	OTIR2020	VITA	PIETRE	PENTA	POLIS	NANOXM	CENTO	PIERRE	POLO12	POLITER
OPTOSCANA	**19**		6	17	1	9	13	14	5	14	11	11
INNOPAPER	1	**4**	3	4	5	4	3	3	4	5	2	5
OTIR2020	1		**10**		1	2			4	1	2	
VITA	9		4	**10**	3	5	6	7	9	9	7	4
PIETRE												
PENTA					1	**1**					2	
POLIS							**2**			1		
NANOXM	2		1	2		2	1	**4**	3	2	1	2
CENTO												
PIERRE			1	4	1	1	3	1	1	**3**	5	2
POLO12			1	3	1	1			1		**12**	2
POLITER	13	3	9	15	2	14	12	8	14	13	20	**22**

Grafo 9 Reti di relazioni tra i poli attraverso i link nelle pagine web dei poli, dei soggetti gestori, delle società con cui hanno accordi di collaborazione, delle società partecipanti nei soggetti gestori, dei laboratori e degli incubatori

Legenda: azzurro = fascia 1, rosso = fascia 2, verde = fascia 3. Lo spessore del legame tra i nodi è proporzionale al numero di soggetti in comune. Sono omessi i loop

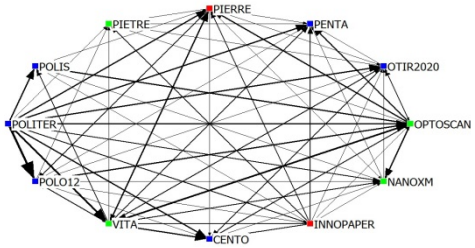


Tabella 11 Misure di centralità della rete unimodale per le reti virtuali

	OutDegree	InDegree	Out Bonacich Power	In Bonacich Power	OutCloseness	InCloseness	Betweenness
OPTOSCANA	11	6	2374.62	1137.04	12	23	2.14
INNOPAPER	12	2	2823.13	274.36	11	28	0
OTIR2020	7	8	1110.43	1716.98	17	19	1.98
VITA	11	7	2374.62	1443.81	12	20	3.98
PIETRE	0	8	0	1863.2	44	16	0
PENTA	3	9	237.41	2041.6	25	18	0.14
POLIS	2	7	378.6	1331.2	23	21	0
NANOXM	10	6	2373.62	1119.19	13	22	0.5
CENTO	0	8	0	1716.97	44	16	0
PIERRE	10	8	1997.21	1604.36	13	20	12.14
POLO12	7	9	1241.35	2041.6	16	18	10.14
POLITER	12	7	2823.13	1443.81	11	20	11.98

Tabella 12 Numero e componenti delle cliques nella rete unimodale delle reti virtuali

CLIQUE	POLO1	POLO2	POLO3	POLO4
CLIQUE1	VITA	NANOXM	PIERRE	POLITER
CLIQUE2	OPTOSCANA	VITA	NANOXM	POLITER
CLIQUE3	VITA	POLO12	POLITER	

Tabella 13 Clique Participation Scores: proporzione di legami che ogni polo ha con le diverse cliques individuate attraverso Ucinet (reti virtuali)

	CLIQUE1	CLIQUE2	CLIQUE3
OPTOSCANA	0.75	1.00	0.67
INNOPAPER	0.25	0.25	0.33
OTIR2020	0.25	0.25	0.33
VITA	1.00	1.00	1.00
PIETRE	0.00	0.00	0.00
PENTA	0.00	0.00	0.33
POLIS	0.25	0.00	0.00
NANOXM	1.00	1.00	0.67
CENTO	0.00	0.00	0.00
PIERRE	1.00	0.75	0.67
POLO12	0.50	0.50	1.00
POLITER	1.00	1.00	1.00

4. Conclusioni

L'indagine linguistica ha permesso di evidenziare differenti livelli di partecipazione attiva da parte dei diversi poli, attraverso l'analisi della provenienza dei contenuti pubblicati da ogni singolo polo all'interno del proprio sito web. Tale partecipazione è stata considerata centrale ai fini della valutazione delle strategie di promozione dei poli. Attraverso l'analisi dei contenuti pubblicati online è stato possibile verificare quanto essi abbiano contribuito in maniera diretta alla creazione di materiali “innovativi”. Focalizzandosi non tanto sulla quantità dei contenuti pubblicati quanto sulla qualità degli stessi, è infatti possibile tracciare una distinzione – seppur semplicistica – tra siti che “creano” e siti che “riportano” contenuti online. Questa distinzione è stata dunque utilizzata come metro di valutazione dell'operato dei poli, ed è stata tradotta operativamente analizzando la provenienza dei testi contenuti all'interno dei siti web dei poli, al fine di analizzare quanto ogni polo abbia “creato” o “riportato” contenuti. L'analisi ha preso in considerazione 24 termini – scelti in base alla loro relazione con il concetto di “polo d'innovazione” – che sono stati utilizzati come filone d'indagine per identificare quei testi relativi alle attività, alle tecniche di promozione, e ai campi di ricerca dei poli. Attraverso questo esame è stato pertanto possibile evidenziare come - in linea di massima - i siti con la maggiore quantità di contenuti siano siti che riportano contenuti terzi (si vedano ad esempio Polis e Vita); e come i siti con la minore quantità di contenuti siano siti che “creano” nuovi contenuti online (si vedano ad esempio Nanoxm, Otir2020, Innopaper). Per quanto questi risultati possano sembrare contrastanti con il concetto di “presenza online” (i.e. la strategia di promozione ottenuta attraverso la costante pubblicazione di contenuti, che permette di ottenere visibilità grazie al continuo aggiornamento del proprio sito web), ai fini della valutazione abbiam deciso di focalizzarci su aspetti qualitativi piuttosto che su strategie promozionali improntate sullo sfruttamento di algoritmi di ricerca e di ranking. Ciò al fine di valutare il reale impatto che ogni singolo polo ha nell'“innovazione”, qui intesa come “creazione” di materiali originali. Sebbene questa distinzione possa essere considerata riduttiva, la messa a sistema dei risultati dell'analisi linguistica con i risultati dell'analisi relazionale ha permesso di ottenere un quadro valutativo dell'operato dei poli - in termini di “presenza online” – basato sulla qualità dei contenuti e delle relazioni.

Nell'analisi delle reti virtuali abbiamo evidenziato tre aspetti. Il primo riguarda la comunicazione web come misura delle connessioni tra i poli: emerge una scarsissima connessione tra i poli, in termini di citazioni reciproche; non è generalizzata la presenza di citazioni dei soggetti gestori dei poli stessi, ed è molto asimmetrica la citazione dei soggetti

con cui si sono sottoscritti accordi di collaborazione. Nel complesso non sembra che lo spazio virtuale sia stato scelto come canale di rafforzamento della dimensione sistemica dell'azione dei poli. Il secondo aspetto preso in esame è la comunicazione web come misura dello spazio di informazione condiviso: accanto alla presenza più significativa di alcuni dei poli, osserviamo l'assenza di un polo (Innopaper) nelle citazioni da parte di altri poli: quel polo, i suoi gestori, le società partecipanti e le strutture collegate non sono citate da nessun altro polo, mentre quel polo è molto attivo nel citare pagine web di attori che appartengono ad altri poli. Prendendo in considerazione la relazione tra i link completi citati da più poli e la relazione con i poli che li hanno citati è emerso come i poli di terza fascia (Vita, Optoscana e Nanoxm) presentino un più alto numero di riferimenti alle stesse informazioni (link completi di pagine web). Le complementarità tra ambiti tecnologici potrebbero spiegare questo risultato, ma è possibile anche che questi poli abbiano usato maggiormente la comunicazione web proprio per sopperire alla loro ridotta dimensione.

In generale, una delle evidenze che emerge dall'analisi dei siti web è proprio la diversa percezione che i capofila e i soggetti gestori hanno avuto dei poli di innovazione: se da un lato infatti alcuni poli di innovazione sono riusciti ad essere percepiti come un'entità disgiunta dai soggetti gestori e con uno specifico obiettivo e ruolo all'interno del sistema di trasferimento tecnologico toscano, altri sono stati invece concepiti come semplici “progetti” del capofila[10].

[10] Su questo aspetto si veda l'analisi critica proposta da M. Russo, A. Caloffi, F. Rossi, V. Fiordelmondo, S. Ghinoi (2015), Evaluating the performance of innovation intermediaries: insights from the experience of Tuscany’s innovation poles, DEMB Working Paper series n. 62

Appendice 1: Metodologia dell'analisi linguistica

L'analisi linguistica dei siti internet dei poli, e la successiva analisi delle relazioni virtuali, ha richiesto un trattamento specifico dei dati disponibili online. In questo documento vengono riportati gli strumenti utilizzati (1.1), mentre nella sezione 2 vengono riportate le due fasi di trattamento dei dati. In 2.1 sono descritti i passaggi che sono stati eseguiti per la raccolta delle pagine web dei siti dei poli; in 2.2 sono descritti i passaggi che sono stati utilizzati per la raccolta dei link all'interno dei siti dei poli. Infine in 2.3 sono riportati i comandi utilizzati per la raccolta (2.3.1) e la pulizia (2. Ogni sezione comprende una prima parte in cui è riassunto il procedimento in maniera discorsiva, ed una seconda parte in cui sono riportati i comandi utilizzati per eseguire i singoli passaggi – corredati di commenti e note.

Strumenti

Sia la raccolta e successiva pulizia dei dati sia l'estrazione dei link dalle pagine web sono state effettuate su un server Ubuntu 14.10, quad-core, 4GB di RAM, attraverso strumenti da riga di comando. I seguenti strumenti e comandi sono stati utilizzati.

Raccolta e pulizia delle pagine web

httrack: https://www.httrack.com/ Strumento per creare copie locali di siti web.

lynx: http://lynx.browser.org/ Browser web da riga di comando; permette di effettuare il *dump* (la copia dei dati) del testo contenuto all'interno di pagine HTML, eliminando il codice HTML delle pagine.

find: http://www.gnu.org/software/findutils/manual/html_mono/find.html comando che permette di effettuare ricerche sui nomi dei file.

grep: http://www.gnu.org/software/grep/manual/grep.html comando per effettuare ricerche all'interno di file.

perl: https://www.perl.org/ versione per riga di comando, utilizzato per effettuare sostituzioni ed eliminazioni di testo attraverso espressioni regolari su file multipli.

rm: https://www.gnu.org/software/coreutils/manual/html_node/rm-invocation.html#rm-invocation comando per eliminare file.

wc: http://www.linfo.org/wc.html comando per contare il numero di elementi (parole, caratteri, ecc...) all'interno di un file.

shuf: http://www.gnu.org/software/coreutils/manual/html_node/shuf-invocation.html#shuf-invocation comando per estrarre in maniera casuale un numero preciso di elementi da una lista.

Raccolta dei link contenuti nelle pagine web

urifind: http://www.unix.com/man-page/debian/1p/urifind/ script per l'individuazione di link ipertestuali all'interno di pagine html.

sort: https://www.gnu.org/software/coreutils/manual/html_node/sort-invocation.html comando per ordinare il contenuto di file di testo secondo un determinato criterio.

uniq: http://unixhelp.ed.ac.uk/CGI/man-cgi?uniq comando per individuare elementi ripetuti all'interno di file.

Trattamento dei dati

Raccolta e pulizia delle pagine web

Raccolta delle pagine web

Dopo aver raccolto in un file gli indirizzi dei siti dei poli, e dopo averli manualmente controllati per capirne la struttura, le pagine dei siti web sono state scaricate (una cartella per ogni sito) attraverso l'utilizzo di *httrack*. L'operazione, della durata di circa 12 ore, è stata effettuata in data ##/##/##. Successivamente in data ##/##/## i siti di innopaper e polis-toscana sono stati scaricati nuovamente; questa operazione si è resa necessaria per due motivi. Il sito di innopaper risultava errato, siccome *httrack* aveva provveduto a scaricare solamente le pagine in inglese. Si è quindi proceduto alla raccolta manuale di tutte le pagine in Italiano del sito. Il sito di polis-toscana era stato scaricato in maniera incompleta da *httrack*, a causa di alcune restrizioni da parte del sito stesso. Un differente comando di *httrack* è quindi stato utilizzato.

Pulizia delle pagine web

Una volta copiate le pagine web dei singoli siti, divise in 11 cartelle (una per ogni sito), si è provveduto all'individuazione di quegli elementi (codice sorgente, script, lettori video, ecc...) da escludere dall'analisi linguistica. Ognuno degli 11 siti ha richiesto una strategia diversa; gli 11 procedimenti sono elencati nel dettaglio nella sezione comani (2.3.2). Una volta eliminati gli elementi inutili ai fini dell'analisi linguistica, i file sono stati salvati come file di testo (estensione .txt) e successivamente copiati all'interno di 11 cartelle (una per ogni sito).

Raccolta dei link contenuti nelle pagine web

I link contenuti all'interno dei siti dei poli sono stati estratti dai file html originali scaricati durante il procedimento descritto in 2.1.1. Lo script *urifind* è stato utilizzato all'interno di ognuna delle 11 cartelle, e l'output è stato salvato all'interno di un file di testo, successivamente utilizzato per eliminare le occorrenze dei link interni al sito. Una volta ottenuti 11 file contenenti i link esterni presenti sui siti dei poli, gli indirizzi sono stati ordinati alfabeticamente e contati, in modo da ottenere la frequenza di ognuno di essi.

Comandi

Di seguito vengono riportati i comandi utilizzati per le tre differenti operazioni descritte in precedenza (2.1.1, 2.1.2, 2.2). L'ordine in cui sono presentati rispecchia l'ordine in cui sono stati eseguiti; all'interno dell'elenco dei comandi sono inserite delle note che spiegano l'operazione effettuata.

Raccolta delle pagine web

L'elenco dei siti da scaricare è contenuto nel file "elenco_siti_poli.in_percrawl.txt", che contiene quanto segue:

http://optoscana.net/
http://innopaper.lucense.it/
http://www.otir2020.it/
http://www.scienzedellavita.it/
http://www.polopenta.it/
http://www.polis-toscana.it/
http://www.nanoxm.it/
http://polocento.it/
http://www.polopierre.it/
http://www.polo12.it/
http://www.distrettoict-robotica.it/

Le pagine sono state scaricate con:

```
httrack -q -%i -w -%L
/home/elenco_siti_poli.in_percrawl.txt -O
"/home/websites/poli.in" -%P -N0 -s2 -o0 -x -p7 -D -a -
K0 -c10 -%k -T -A500000 -F "Mozilla/4.5 (compatible;
HTTrack 3.0x; Windows 98)" -*.png -*.gif -*.jpg -*.png
-*.css -*.js -*.pdf -ad.doubleclick.net/* -%s -%u
```

Pulizia delle pagine web

Le pagine html sono prima di tutto state convertite in file di testo (.txt) utilizzando il seguente script, adattato di volta in volta alla cartella in cui veniva eseguito:

```
#!/bin/sh
for file in `find . -iname \*.html`
do
new=`basename $file html`
lynx -dump -nolist $file > /DIRECTORY/${new}txt
done
```

Questo script è stato usato direttamente solo su 4 cartelle (innopaper.lucense.it, www.otir2020.it, www.polis-toscana.it, www.scienzedellavita.it), siccome le rimanenti 7 contengono sottodirectory dentro a cui ci sono ulteriori file .html. Per le rimanenti 7 è stato necessario copiare tutti i file .html delle sottodirectory in un'unica cartella; siccome molti file .html in sottocartelle differenti hanno lo stesso nome, copiare semplicemente i file .html in un'unica cartella vuole dire sovrascrivere i diversi file con lo stesso nome. Ho quindi utilizzato il seguente comando per aggiungere, tra il nome del file e l'estensione, un suffisso contenente il path del file:

```
find . -type f -name '*.html' -exec bash -c 'ext="${0##*.}"; base="$(basename "$0" ."${ext}")"; dirs="$(dirname "$0" | tr / _)"; new="/DIRECTORY/${base}_${dirs}.${ext}"; mv "$0" "${new}"' {} \;
```

A questo punto è stato utilizzato lo script di lynx illustrato precedentemente sulle nuove cartelle contenenti tutti i file .html in un'unica cartella.

Una volta ottenuta la versione in formato testo di tutte le pagine degli 11 poli, differenti strategie sono state adottate per pulire il testo non necessario all'analisi linguistica. Di seguito sono riportati i passaggi, suddivisi per nome del polo. Nota bene: nella descrizione dei comandi viene usato il termine “testa” per riferirsi a quella parte di testo da eliminare che appare prima dell'inizio del testo da preservare; viene usato il termine “coda” per riferirsi a quella parte di testo da eliminare che appare dopo la fine del testo da preservare.

innopaper.lucense.it

Una volta convertite con lynx i 29 file html, viene pulita la testa che presenta sempre la dicitura "home page", fino al link "* Contatti" con:

```
perl -0777 -p -i -e "s/Home page\n.*?\*
Contatti//s" *.txt
```

ora tolgo la coda (da "Torna indietro" e poi il disclaimer) con:

```
perl -0777 -p -i -e "s/Torna
indietro\n{1,3}.*?Mappa del sito.*//s" *.txt
```

In ultimo viene eliminato un file che contiene testo interamente in inglese:

```
rm Le_attività_di_R&S_in_Europa_-_Innopaper.txt
```

optoscana.net

Prima ancora di convertire le pagine html in file di testo (con lo script di *lynx*), vengono tutto eliminati i file che contengono unicamente dei redirect:

```
find . -exec grep -l '<TITLE>Page has
moved</TITLE>' {} \;| xargs rm
```

Tra i file txt ce ne sono alcuni con solo testo in inglese, e si differenziano per la dicitura "en" nel nome. Vengono quindi eliminati i file (125 in tutto) che seguono la struttura delle pagine in inglese con

```
rm index_._en*
```

e viene anche eliminato il file en_._.txt (sempre in inglese).

Una serie di file contiene link a feeds (43 in tutto). Questi hanno come suffisso "...feed.txt". Vengono eliminati:

```
rm *_feed.txt
```

Viene quindi eliminato il disclaimer in fondo alle pagine

```
perl -0777 -p -i -e 's/Il Polo Optoscana \|
Optoelettronica.*//s' *.txt
```

e i file che presentano la struttura index[Numero][Numero][Numero][Numero].txt perché contengono redirect.

Tolgo la testa con:

```
perl -0777 -p -i -e "s/\#Optoscana \» Feed
Optoscana.*?Newer posts \→//s" *.txt
perl -0777 -p -i -e "s/\#Optoscana \» Feed
Optoscana.*?\← Older posts//s" *.txt
perl -0777 -p -i -e "s/\#Optoscana \» Feed
Optoscana.*?Next \→//s" *.txt
perl -0777 -p -i -e "s/\#Optoscana \» Feed
Optoscana.*?\* CONTATTI//s" *.txt
```

e la coda con:

```
perl -0777 -p -i -e "s/Comments are closed.*//s"
*.txt
```

```
perl -0777 -p -i -e "s/Post navigation.*//s"
*.txt
perl -0777 -p -i -e "s/\* NETWORK.*//s" *.txt
perl -0777 -p -i -e "s/This entry was posted in
Uncategorized.*//s" *.txt
```

polocento.it

Tra le poche pagine (247) alcune contengono informazioni non utili (link che si riferiscono alla struttura del sito, disclaimer, breve descrizione del polo presente in ogni pagina) che vengono eliminate:

```
perl -0777 -p -i -e 's/#Polo Cento.*?o
Consulenze//s' *.txt
perl -0777 -p -i -e 's/POLO DI COMPETENZA PER IL
SISTEMA INTERNI.*//s' *.txt
```

www.distrettoict-robotica.it

Vengono prima di tutto eliminati i file con testo unicamente in inglese, che presentano nel nome la dicitura "_en_" (488 in totale):

```
find . -iname '*_en_*' | xargs rm
```

altri file sempre in inglese invece hanno la dicitura "..._en.txt" (33 in totale), e vengono eliminati con

```
find . -iname '*_en.txt' | xargs rm
```

Nei rimanenti file ci sono in testa e in coda i) link fb, etc.. e mappa del sito e ii) disclaimer. Una prima parte della coda viene eliminata con:

```
perl -0777 -p -i -e 's/Operazione finanziata nel
quadro del POR FESR.*//s' *.txt
```

Un'altra parte di coda viene eliminata con

```
perl -0777 -p -i -e 's/AREA RISERVATA   Nome
utente.*//s' *.txt
```

L'ultima parte di coda viene eliminata con:

```
perl -0777 -p -i -e 's/Soggetti Promotori.*?\*
cnr.jpg.*//s' *.txt
```

La testa viene eliminata con:

```
perl -0777 -p -i -e 's/fb.*?\* Contatti//s' *.txt
```

www.nanoxm.it

Viene eliminata la testa (menu e mapsite) e la coda (disclaimer) con

Testa:

```
perl -0777 -p -i -e 's/#nanoxm ».*?\+ Eventi//s'
*.txt
```

Coda:

```
perl -0777 -p -i -e 's/Dove siamo.*//s' *.txt
```

www.otir2020.it

Alcuni file, nonostante il suffisso .txt, sono in realtà dei file binari. Riportano la dicitura imghandler* nel nome, e vengono eliminati (28 in totale) con:

```
rm imghandler*
```

Ai rimanenti file vengono eliminate testa e coda. Per la testa:

```
perl -0777 -p -i -e "s/Officina toscana per
l\'innovazione e la ricerca di.*?\* Contatti//s"
*.txt
```

per la coda:

```
perl -0777 -p -i -e "s/OTIR 2020 \- Operazione
finanziata nel quadro.*//s" *.txt
```

www.polis-toscana.it

A causa della grande varietà di script e moduli presenti nelle pagine html, la conversione diretta in formato di testo con *lynx* risulta non valida per preservare il testo per l'analisi linguistica. E' stato quindi necessario pulire prima i file html, convertirli in txt e successivamente pulire i txt.

La testa dei file viene eliminata con:

```
perl -0777 -p -i -e 's/.*?<head>/<head>/s' *.html
```

Per eliminare il sorgente in coda viene usato

```
perl -0777 -p -i -e 's/iCagenda by Jooml.*//s'
*.html
```

I file html puliti vengono quindi convertiti in txt con lo script di *lynx*.

Viene tolta la parte di testo relativa alla registrazione con

```
perl -0777 -p -i -e "s/Registrazione consentita
ai soli associati a
Polis.*?\/controller\/view.php//s" *.txt
```

I file presentano ancora molte sezioni di testo da eliminare: i successivi passaggi sono quindi:

I) eliminazione del nome dei link in testa (home, contatti), intestazione, e i link sotto (home, governance polis...).

II) eliminazione delle diciture "News Polis", "In evidenza" e "News Finanziamenti", ma non i link relativi a queste sezioni, perché sono dinamici e riportano i titoli degli articoli. Vengono inoltre eliminati gli asterischi davanti ad ognuno dei titoli degli articoli.

III) Eliminazione di tutta la sezione "Il polo" con relativi sottolink.

IV) Eliminazione della sezione "Distretto" con relativi sottolink.

V) Eliminazione della sezione "Finanziamenti" con relativi sottolink.

VI) Eliminazione di ciò che rimane della sezione "Registrazione" (ovvero il titolo e il codice "on line 0")

VII) Eliminazione della sezione "Documenti" con relativi sottolink.

VIII) Eliminazione della dicitura "Calendario Eventi" che è una rimanenza della precedente pulizia dell'area dell'agenda di Joomla.

IX) Eliminazione degli asterischi che sono rimasti di fronte agli articoli della sezione II).

Di seguito i comandi usati per ognuna delle operazioni riportate sopra (I-IX)

I) e II):

```
perl -0777 -p -i -e "s/\* Home.*?News Polis\n//s"
*.txt
```

II) altre diciture da togliere

```
perl -0777 -p -i -e "s/In evidenza\n//s" *.txt
perl -0777 -p -i -e "s/In evidenza\n//s" *.txt
perl -0777 -p -i -e "s/News Finanziamenti\n//s"
```

```
*.txt
```

III):

```
perl -0777 -p -i -e "s/Il Polo\n.*?\* Polis in
cifre\n//s" *.txt
```

IV):

```
perl -0777 -p -i -e "s/Distretto\n.*?\*
Finanziamenti\n//s" *.txt
```

V):

```
perl -0777 -p -i -e "s/Finanziamenti\n.*?\*
Nazionali e UE\n//s" *.txt
```

VI):

```
perl -0777 -p -i -e "s/Registrazione\n.*?on line
0\n//s" *.txt
```

VII):

```
perl -0777 -p -i -e "s/Documenti\n.*?\*
Photogallery\n//s" *.txt
```

VIII):

```
perl -0777 -p -i -e "s/Calendario Eventi//s"
*.txt
```

IX):

```
perl -0777 -p -i -e "s/  \* //" *.txt
```

Circa 1100 file di testo sono in realtà dei file binari, che vengono eliminati con:

```
find . -exec grep -l 'Strict Standards' {} \;|
xargs rm
```

I rimanenti file di testo presentano ancora tracce di script, che vengono eliminate con:

```
perl -0777 -p -i -e "s/ Warning\: Illegal string
offset.*//s" *.txt
```

Rimangono inoltre parti di testo che si riferiscono a feed RSS, funzioni web, titoli dei link, e che vengono eliminate con:

```
perl -0777 -p -i -e "s/\#RSS 2\.0 Atom 1\.0//s"
*.txt
perl -0777 -p -i -e
"s/\*\*\|.*?Print.*?\|\*\*//s" *.txt
perl -0777 -p -i -e "s/\*\*Articles\*\*//s" *.txt
perl -0777 -p -i -e "s/\* \w.*//" *.txt
```

Nonostante la pulizia dei dati, il software per l'analisi linguistica non riesce a caricare il corpus siccome i file sono troppi. Viene dunque deciso di creare un corpus più piccolo, contenente un quinto dei file originali. I testi da includere in questa versione ridotta vengono scelti casualmente con il seguente comando:

```
shuf -zen13531 * | xargs -0 cp -t
/root/poliin/TXT/www.polis-toscana.itRANDOM13531
```

www.polo12.it

Dai file di testo viene innanzitutto eliminata la coda relativa a Joomla:

```
perl -0777 -p -i -e "s/(?s)Copyright 2008 \- 2011
JoomlaShine\.com.*?Joomla templates by
Joomlashine//s" *.txt
```

Viene poi eliminata la sitemap:

```
perl -0777 -p -i -e "s/(?s)\* Home\n.*?\* Il
Polo.*?\* News//s" *.txt
```

4072 file sono pagine relative a contatti email, e presentano la dicitura "mailto " prima dell'estensione del file. I file vengono eliminati con:

```
rm *mailto.txt
```

Viene poi eliminata la sezione del "Calendario" in coda con:

```
perl -0777 -p -i -e
"s/Calendario\n{1,4}\s{1,30}\«.*//s" *.txt
```

ed altri elementi in coda con:

```
perl -0777 -p -i -e "s/Ultimi eventi.*//s" *.txt
```

Rimangono altre stringhe di testo che vengono eliminate con:

```
perl -0777 -p -i -e "s/\#RSS 2\.0 Atom 1\.0//s"
*.txt
perl -0777 -p -i -e "s/\* Print//s" *.txt
perl -0777 -p -i -e "s/\* EMAIL//s" *.txt
```

Viene dunque eliminata la sezione “News”, che riporta titoli già presenti nelle singole pagine:

```
perl -0777 -p -i -e
"s/News\n.*?\*.*?HomeEventi//s" *.txt
```

Vengono successivamente eliminate le sezioni di testo create dall'applicazione “calendario”:

```
perl -0777 -p -i -e
"s/(01|02|03|04|05|06|07|08|09|10|11|12|13|14|15|
16|17|18|19|20|21|22|23|24|25|26|27|28|29|30|31)
(January|February|March|April|May|June|July|Augus
t|September|October|November|December) 20[0-9][0-
9] -
(01|02|03|04|05|06|07|08|09|10|11|12|13|14|15|16|
17|18|19|20|21|22|23|24|25|26|27|28|29|30|31)
(January|February|March|April|May|June|July|Augus
t|September|October|November|December) 20[0-9][0-
9]//" *.txt
perl -0777 -p -i -e
"s/(Monday|Tuesday|Wednesday|Thursday|Friday|Satu
rday|Sunday)\n(01|02|03|04|05|06|07|08|09|10|11|1
2|13|14|15|16|17|18|19|20|21|22|23|24|25|26|27|28
|29|30|31)
(January|February|March|April|May|June|July|Augus
t|September|October|November|December)//s" *.txt
perl -0777 -p -i -e
"s/\[(January|February|March|April|May|June|July|
August|September|October|November|December) 20[0-
9][0-9].*?\]//" *.txt
perl -0777 -p -i -e "s/By Month//" *.txt
perl -0777 -p -i -e "s/By Week//" *.txt
perl -0777 -p -i -e "s/By Day//" *.txt
perl -0777 -p -i -e "s/Events for the
(day|week|month) \://" *.txt
perl -0777 -p -i -e "s/Events Calendar//" *.txt
perl -0777 -p -i -e "s/\* There are no events on
this date//" *.txt
perl -0777 -p -i -e
"s/(Monday|Tuesday|Wednesday|Thursday|Friday|Satu
rday|Sunday)\n\h{2,1000}(01|02|03|04|05|06|07|08|
09|10|11|12|13|14|15|16|17|18|19|20|21|22|23|24|2
5|26|27|28|29|30|31)
(January|February|March|April|May|June|July|Augus
t|September|October|November|December)//s" *.txt
perl -0777 -p -i -e "s/Events for//" *.txt
perl -0777 -p -i -e
"s/(Monday|Tuesday|Wednesday|Thursday|Friday|Satu
rday|Sunday)
(01|02|03|04|05|06|07|08|09|10|11|12|13|14|15|16|
```

```
17|18|19|20|21|22|23|24|25|26|27|28|29|30|31)
(January|February|March|April|May|June|July|Augus
t|September|October|November|December) 20[0-9][0-
9]//" *.txt
perl -0777 -p -i -e "s/\* No events//" *.txt
perl -0777 -p -i -e "s/\Close//" *.txt
```

www.polopenta.it

Alcuni file binari "mascherati" da file di testo vengono eliminati manualmente. Successivamente viene eliminata la testa (sitemap) con:

```
perl -0777 -p -i -e "s/(?s)#start.*?\* contatti
\>//s" *.txt
```

www.polopierre.it

Viene eliminata la testa con:

```
perl -0777 -p -i -e "s/#Polo Pierre.*?\*
Contatti//s" *.txt
```

e la coda con:

```
perl -0777 -p -i -e "s/search\.\.\..*//s" *.txt
```

www.scienzedellavita.it

In coda ai file presente una sezione (“applicazione”), che appare anche in inglese ("apply"). Queste code vengono eliminate con:

```
perl -0777 -p -i -e "s/Apply\n.*//s" *.txt
perl -0777 -p -i -e "s/Applica\n.*//s" *.txt
```

Viene quindi eliminata la sitemap:

```
perl -0777 -p -i -e "s/Menu principale\n.*?\*
Contatti//s" *.txt
```

e successivamente la testa (in Italiano):

```
perl -0777 -p -i -e "s/Salta al contenuto
principale.*?\* ENG//s" *.txt
```

e la testa in Inglese con:

```
perl -0777 -p -i -e "s/Skip to main content.*?\*
ENG//s" *.txt
```

Appendice 2: Principali domini citati

tipologia	Denominazione	ranking (n. di poli in cui sono richiamati)	nome dominio	poli in cui è richiamato nel sito web							
altro	Regione Toscana	8	http://www.regione.toscana.it/	OPTOSCANA	OTIR2020	PENTA	PIERRE	POLIS	POLITER	POLO12	VITA
altro	Commisione Europea	5	http://ec.europa.eu/	NANOXM	OTIR2020	POLIS	POLITER	VITA			
altro	Sviluppo Toscana	5	http://www.sviluppo.toscana.it/	OTIR2020	PIERRE	POLITER	POLO12	VITA			
altro	Toscana Promozione	4	http://adesioni.toscanapromozione.it/	OPTOSCANA	PIERRE	POLITER	POLO12				
centro servizi	Apre Toscana	4	http://www.apretoscana.org/	NANOXM	PIERRE	POLITER	VITA				
universita'	Fondazione ricerca e innovazione	4	http://www.fondazionericerca.unifi.it/	OPTOSCANA	PIERRE	POLIS	POLITER				
universita'	UNIPI	4	http://www.unipi.it/	OPTOSCANA	INNOPAPER	POLITER	VITA				
centro pubblico di ricerca	Apre	3	http://www.apre.it/	NANOXM	POLITER	VITA					
altro	B2Match	3	http://www.b2match.eu/	OPTOSCANA	POLO12	VITA					
centro pubblico di ricerca	CNIT	3	http://www.cnit.it/	OPTOSCANA	POLITER	VITA					
centro servizi	I2T3	3	http://www.i2t3.unifi.it/	OPTOSCANA	OTIR2020	POLO12					
centro pubblico di ricerca	INSTM	3	http://www.instm.it/	INNOPAPER	NANOXM	VITA					
impresa	Kayser	3	http://www.kayser.it/	OPTOSCANA	POLITER	VITA					
centro pubblico di ricerca	LENS	3	http://www.lens.unifi.it/	OPTOSCANA	POLITER	VITA					
centro servizi	Pont Tech	3	http://www.pont-tech.it/	PIERRE	POLITER	POLO12					
altro	Smau	3	http://www.smau.it/	PIERRE	POLITER	VITA					
universita'	Scuola Superiore Sant'anna di Pisa	3	http://www.sssup.it/	OPTOSCANA	POLITER	VITA					
centro privato di ricerca	NEXT	3	http://www.tecnotex.it/	OPTOSCANA	OTIR2020	POLITER					
universita'	UNISI	3	http://www.unisi.it/	NANOXM	OPTOSCANA	POLITER					