Strumenti Qualitativi e Quantitativi di indagine: l'Approccio dell'Analisi Automatica dei Testi

Marco Vassallo, CREA-PB

4/5/6 Luglio 2022, Cogne, Valle d'Aosta

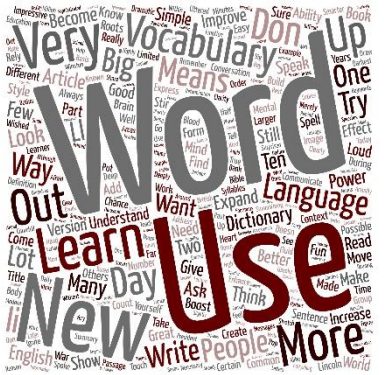
Laboratorio SMART VILLAGE:

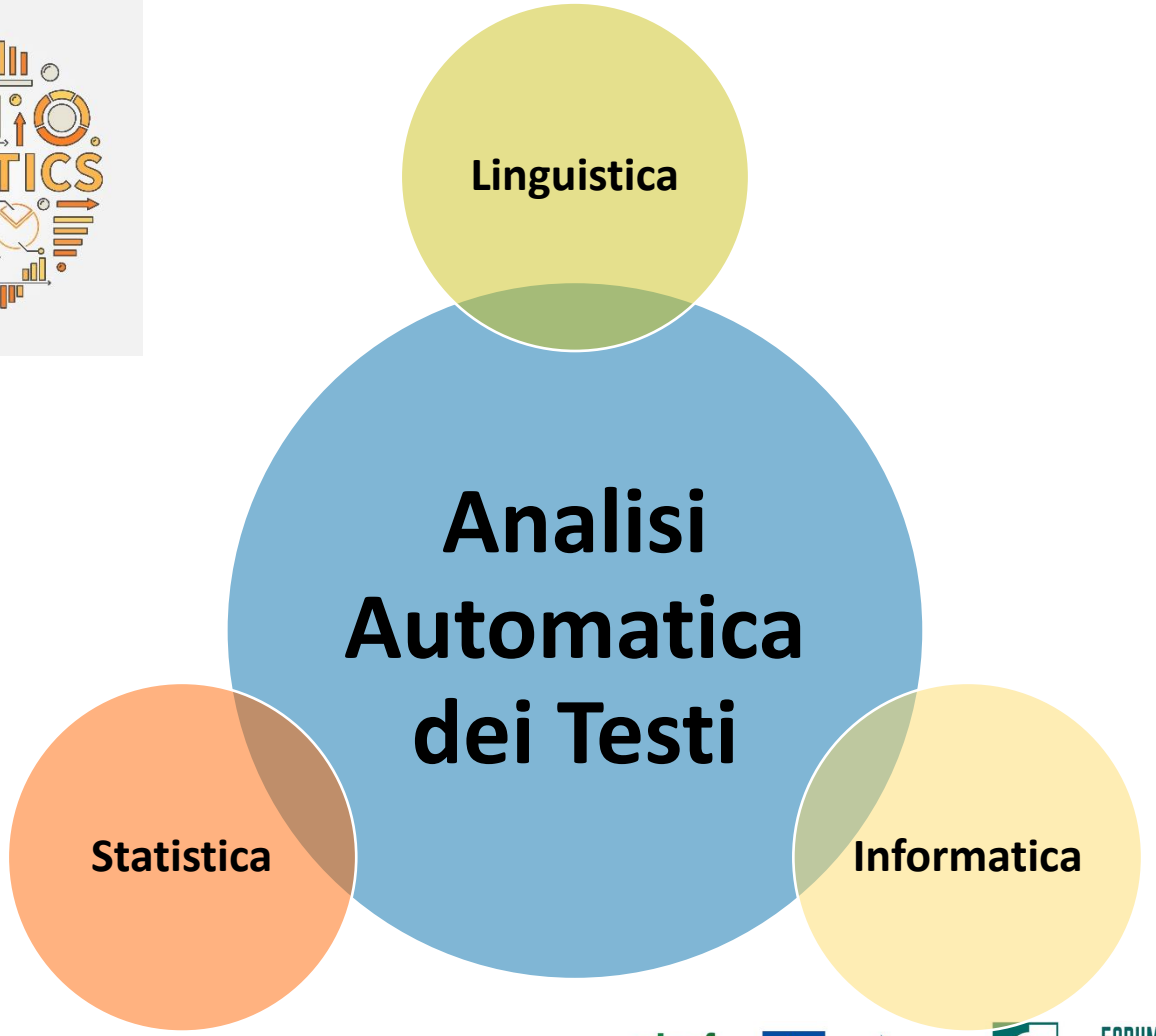
Dalla teoria alla pratica

Sommario

- Analisi Automatica dei Testi: cos'è? Perchè si fa?
- Prerequisiti: questionari strutturati a domande chiuse e aperte – Focus Groups - Interviste in profondità –
- Potenziali applicazioni nell'Agroalimentare
- Attività di Ricerca e Prospettive Future

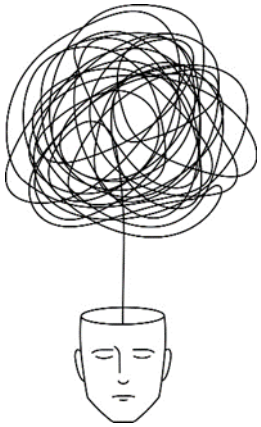
Cos'è l'Analisi Automatica dei Testi?





Si, ma perché si fa?

Già nel 1995 un rapporto della Forrester Research* stimava che l'80% delle informazioni in un'azienda era costituita da testi ed il 20% da dati numerici…

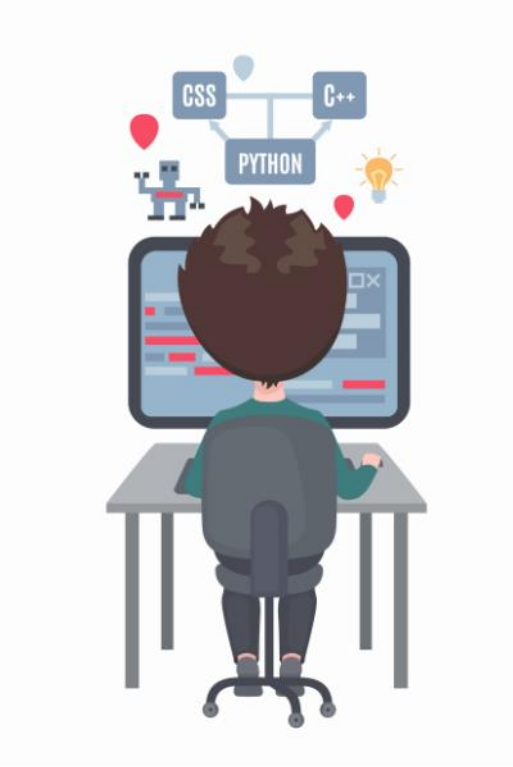


****Bernard Marr (2015) dichiarava su Forbes che nel 2020 ogni essere umano produrrà 1.7 megabyte di dati ogni secondo…**

**Forrester Research (1995). Coping with Complex Data, The Forrester Report, April. In: Giuliano e La Rocca (2008), L'analisi automatica e semi-automatica dei dati testuali. Software e istruzioni per l'uso. Milano, LED, pag. 156.*

***Marr, B. (2015). Big Data: 20 mind boggling facts everyone must read. Forbes (online) 30 settembre. Disponibile su <https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#136378ad17b1>. [Data di accesso: 30 giugno 2022].*

Si, ma che vuol dire analizzare automaticamente un testo?



Testo non strutturato

+

Analisi Lessicale Automatica

→

Testo strutturato

Text Mining

Opinion Mining

Esempio di testo non strutturato: domande aperte.

14. Se non sono state avviate iniziative per la riduzione degli scarti aziendali, quali sono le motivazioni?

……

……

……

15. Sia nel caso siano state avviate iniziative per la riduzione degli scarti, sia nel caso non siano state avviate, può dirci se tali iniziative sono state (o sarebbero, prevedibilmente): *Possibile una sola risposta a domanda,*

15.1. Rispetto all'attuale situazione aziendale (risorse, organizzazione, ecc.):

Insensate		Indifferenti	Sensate	
☐ molto	☐ abbastanza	☐	☐ abbastanza	☐ molto

15.2. Sul piano economico:

Svantaggiose		Indifferenti	Vantaggiose	
☐ molto	☐ abbastanza	☐	☐ abbastanza	☐ molto

15.3. In termini di rilevanza per l'immagine aziendale*:

Non rilevanti		Indifferenti	Rilevanti	
☐ molto	☐ abbastanza	☐	☐ abbastanza	☐ molto

** Considerata la crescente attenzione verso lo spreco da parte dei consumatori.*

Esempio di testo non strutturato: trascrizioni da focus groups – presenza di un moderatore.

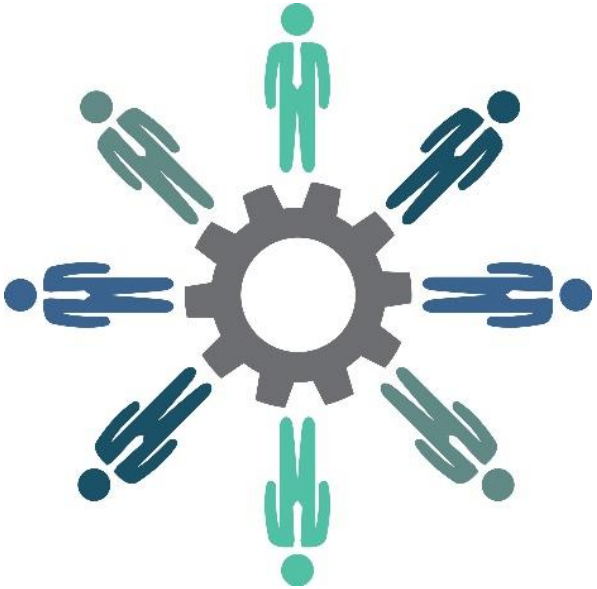
1) Quali vantaggi hanno prodotto per la sua azienda gli aiuti della PAC 2014-2020?

2) Quali novità hanno portato maggior beneficio (es. greening, gestione del rischio, aiuti ai giovani)?

3) E quali sono invece le criticità che ha potuto riscontrare?

4) Alla luce della nuova riforma PAC post 2020 cosa si aspetterebbe in particolare per la sua azienda dalla nuova programmazione?

5) Alla luce dell'emergenza covid 19 cosa si aspetterebbe in particolare per la sua azienda dalla nuova programmazione?



Esempio di testo non strutturato: trascrizioni da interviste in profondità – why questions – Concatenate

1) Quali vantaggi hanno prodotto per la sua azienda gli aiuti della PAC 2014-2020?

2) Quali novità hanno portato maggior beneficio (es. greening, gestione del rischio, aiuti ai giovani)?

3) E quali sono invece le criticità che ha potuto riscontrare?

Risposte:

1) Ha prodotto il vantaggio X -> Perché? Come mai?

2) Ha portato la novità Y -> Perché? Come mai?

3) Ho riscontrato la criticità Z -> Perché? Come mai?

Potenziali applicazioni in Agricoltura

- Focus Group Riso Bio – Text Mining

IRaMuTeQ

R Interface for multidimensional analysis of texts and questionnaires

R & python

http://www.iramuteq.org/

https://www.r-project.org/

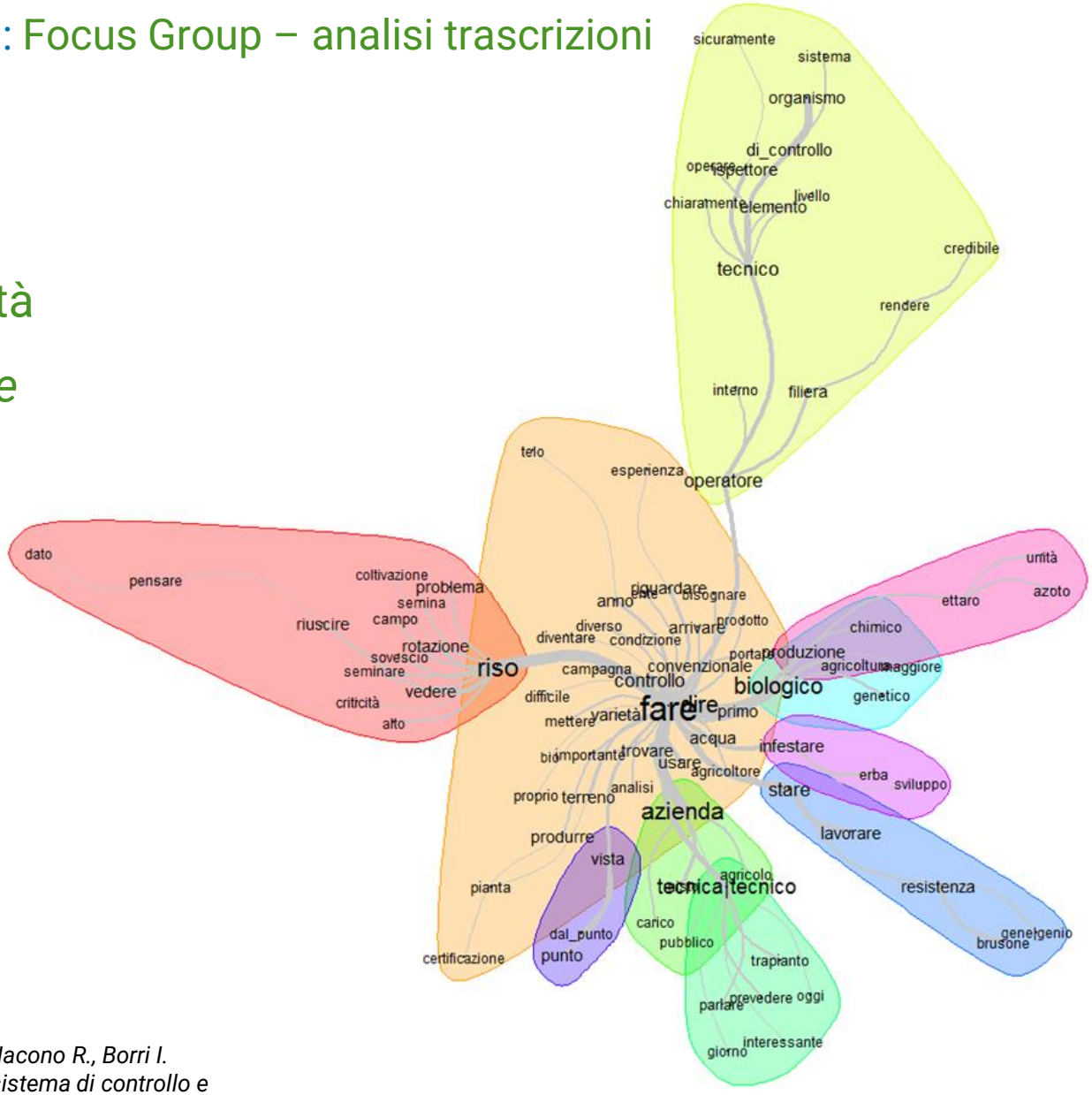
Wordcloud preso da: Monda M., Gabrieli G., & Vassallo M. (2019). Sentiment in Agricoltura: Il termometro dell'Agricoltura. In CREAgritrend: I numeri dell'Agroalimentare Italiano. Bollettino trimestrale edito dal CREA, Centro Politiche e Bio-economia, n°2, I Semestre 2019. ISSN: 2612-6419.

Potenziali applicazioni: Focus Group – analisi trascrizioni

Riso Biologico

Grafo delle Similarità

Co-occurrencies tree

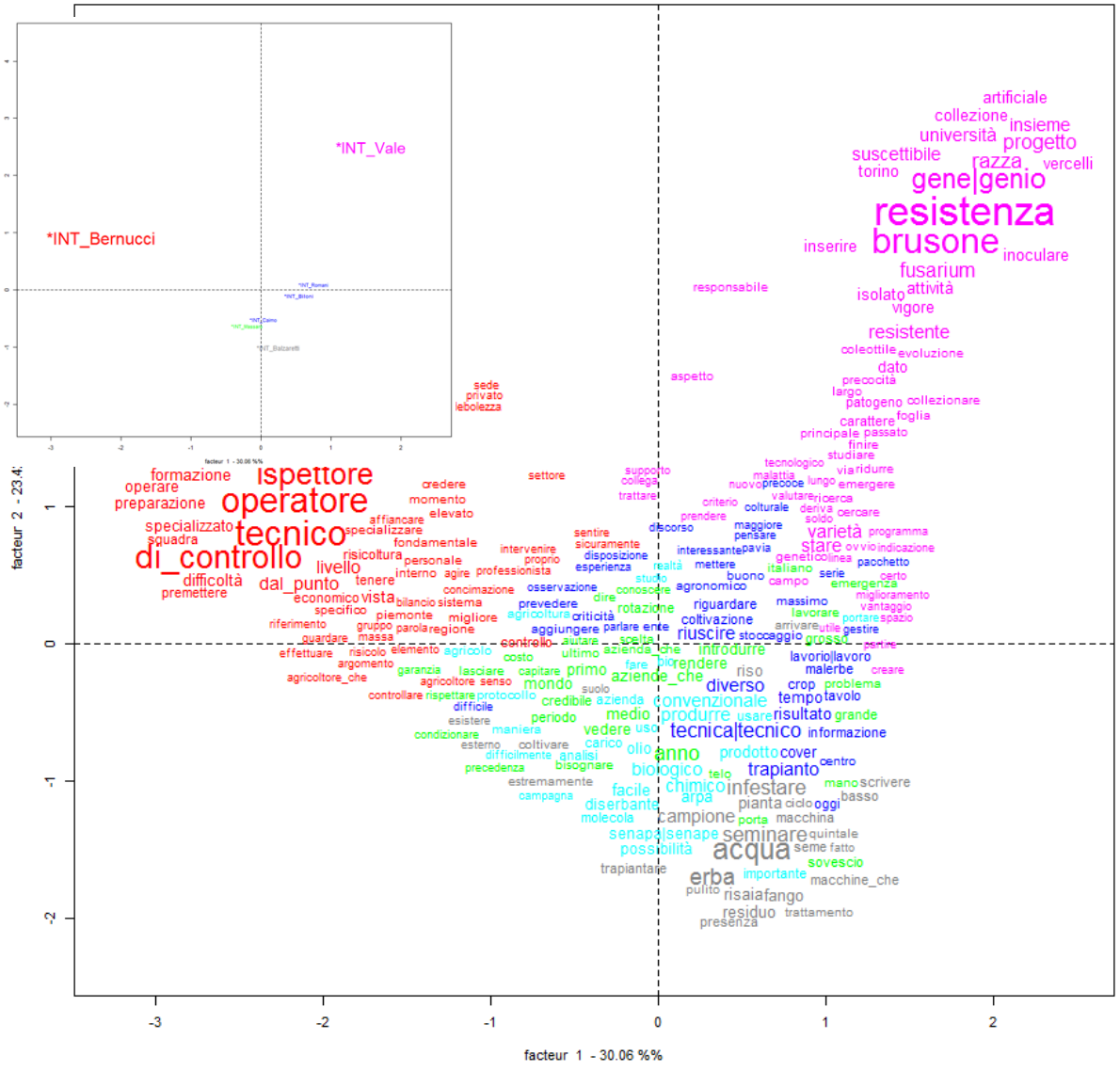


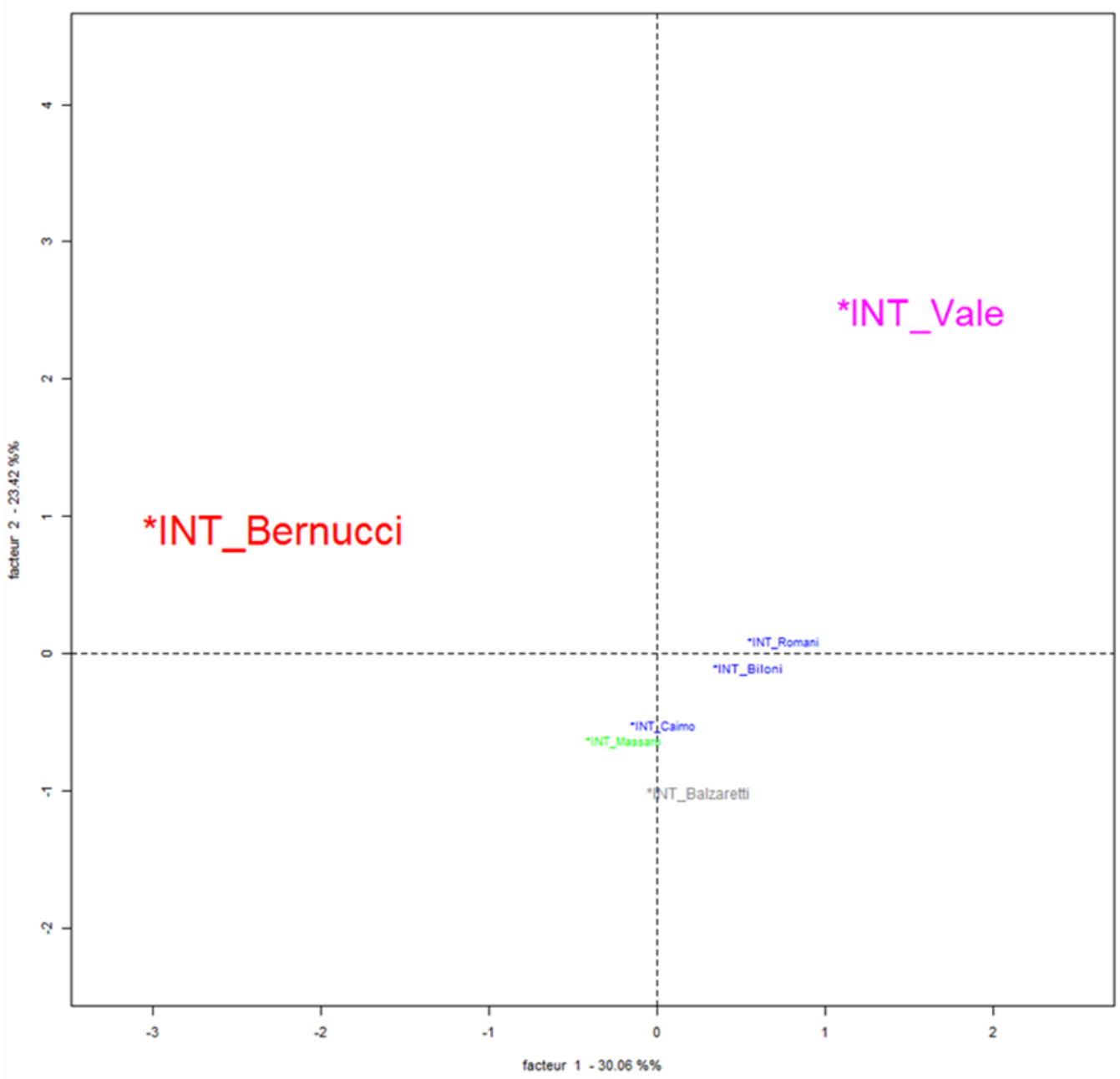
Guccione GD, Vaccaro A., Borsotto P., Vassallo M., Iacono R., Borri I. (2019). Risobiosystems: Documento di analisi del sistema di controllo e certificazione - Report WP3

Potenziali applicazioni: Focus Group – analisi trascrizioni

Riso Biologico

Statistiche descrittive lessicometriche	*Focus Riso*
Numero di testi (intervistati)	7
Numero di occorrenze (N)	13174
Numero di forme grafiche (V)	1937
Type/token ratio = (V/N)*100	14.7%
Frequenza media = N/V	6.80
G di Guiraud	16.87
Coefficiente α	1.25
Numero di hapax	934
Hapax *% sulle occorrenze (N)* *sulle forme (V)*	7.1% 48.2%




*INT_Vale
*INT_Bernucci
*INT_Romani
*INT_Biloni
*INT_Caimo
*INT_Massari
*INT_Balzaretti
facteur 2 - 23.42 %%
facteur 1 - 30.06 %%
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2

Potenziali applicazioni: Focus Group – analisi trascrizioni

Riso Biologico

+-+-+-+-+-+-+-+-+
|i|R|a|M|u|T|e|Q|
+-+-+-+-+-+-+-+-+
Number of texts: 7
Number of text segments: 361
Number of forms: 2679
Number of occurrences: 13174
Number of lemmas: 1937
Number of active forms: 1585
Number of supplementary forms: 310
Number of active forms with a frequency >= 3: 494
Mean of forms by segment: 36.493075
Number of clusters: 6
292 segments classified on 361 (80.89%)

Dendrogramma

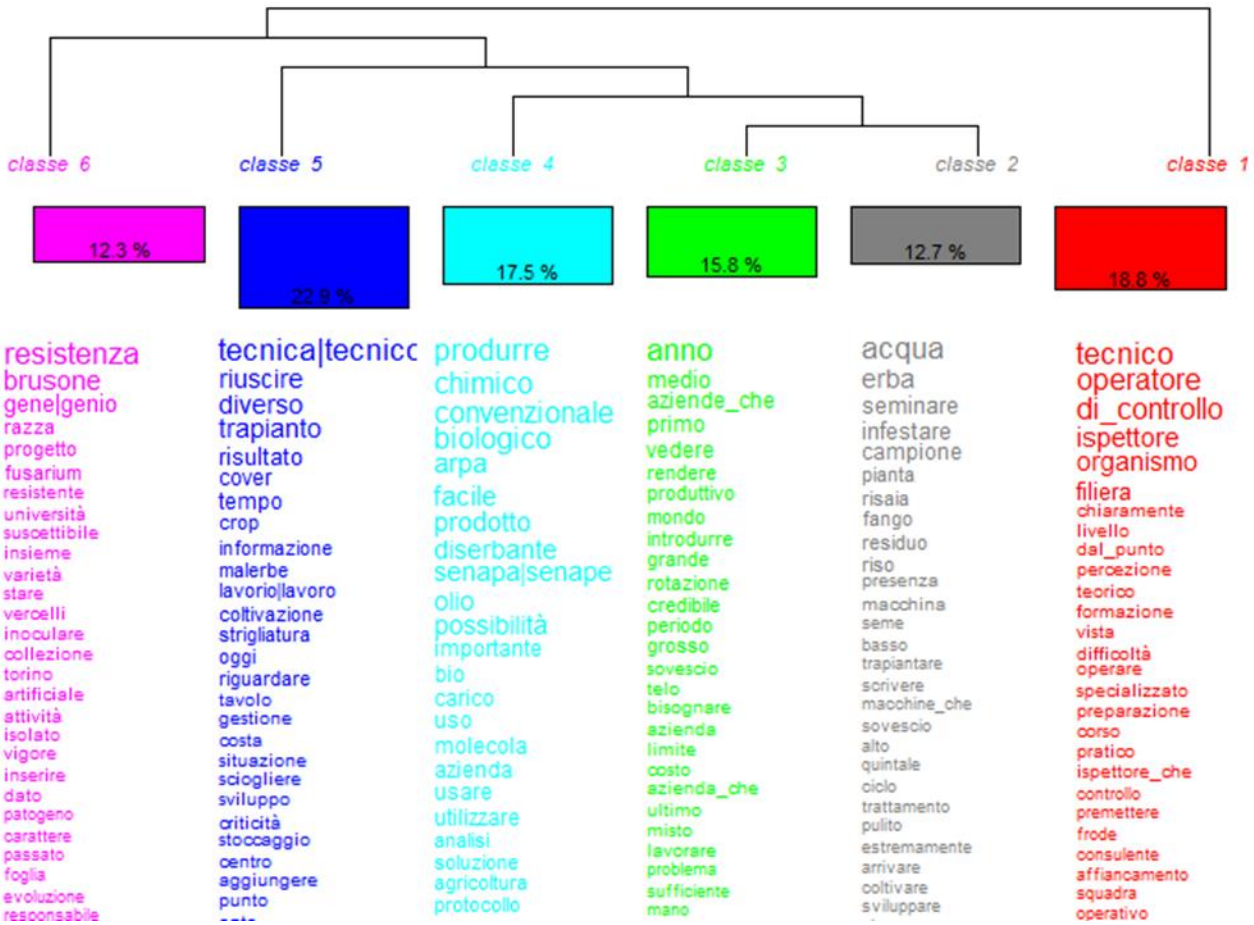
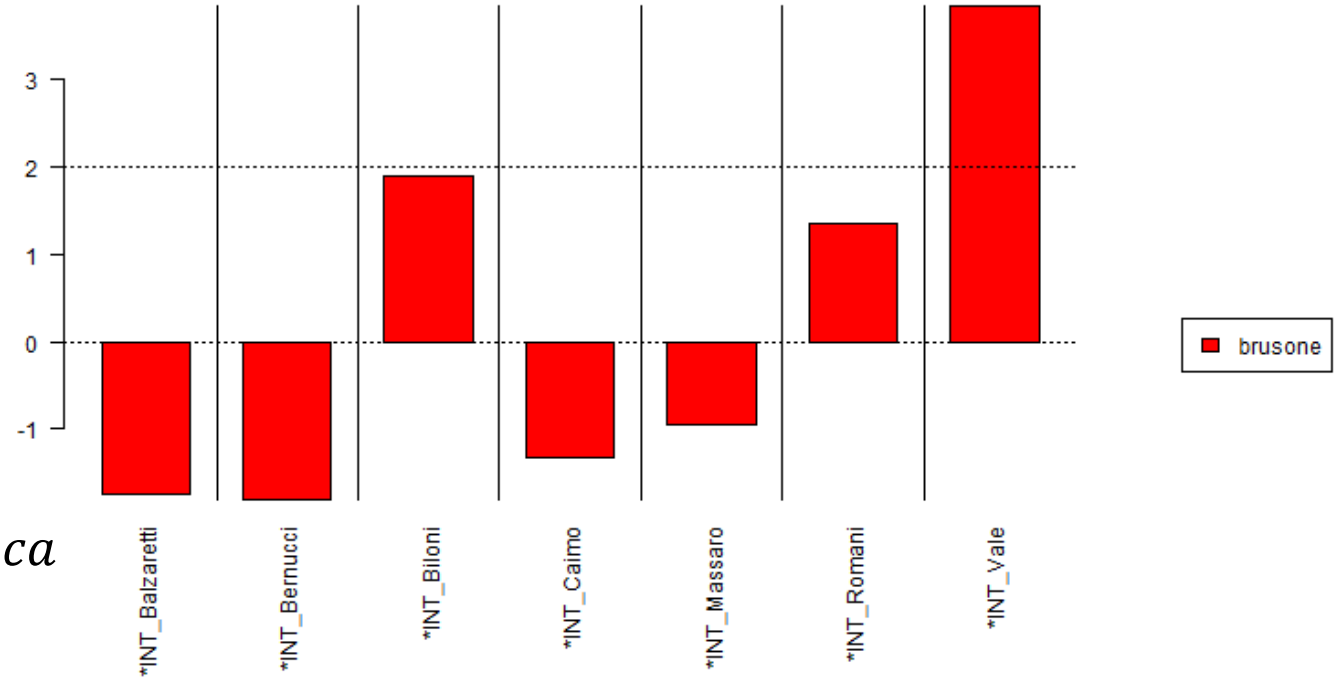


Grafico delle Specificità/Peculiarità per la parola 'brusone'

Riso Biologico

$$^{*}\, z_i = \frac{f_i - f_i^{rif}}{\sqrt{f_i^{rif}}}$$



$z_i > 2\,;\, caratteristica$

$z_i < -2\,;\, rara$

$z_i \approx 0\,;\, banale$

*Bolasco, S. (1999). Analisi Multidimensionale dei dati. Metodi, strategie e criteri di interpretazione. Roma, Carocci.

Potenziali applicazioni in Agricoltura

- **LA VALORIZZAZIONE DEI SERVIZI ECOSISTEMICI PER LO SVILUPPO LOCALE DELLE ZONE RURALI: PROPOSTA DI UN CASO STUDIO PER IL LEADER**

- **IL PROGETTO OROBIKEANDO PROMOSSO DAI GAL VALTELLINA, COLLI DI BERGAMO, VAL BREMBANA 2020, VAL SERIANA E LAGHI BERGAMASCHI**

- **TEXT MINING**

IRaMuTeQ

R Interface for multidimensional analysis of texts and questionnaires

http://www.iramuteq.org/

https://www.r-project.org/

Potenziali applicazioni: Valorizzazione dei servizi ecosistemici

- **IL PROGETTO OROBIKEANDO NASCE DALL'ESIGENZA DI RICLASSIFICARE LE STRADE, A PARTIRE DA QUELLE A BASSA FREQUENZA VEICOLARE, IN DIREZIONE DI UN ASSETTO PROMISCUO ANCHE DI QUELLE PIÙ TRAFFICATE, MA AD ALTA INTENSITÀ CICLISTICA E AD ALTA VALENZA TURISTICA.**

GAL	Unità territorio	Codifica
Valtellina Valle dei Sapori	Alta montagna	VT1
	Media montagna	VT2
	Fondovalle	VT3
Colli di Bergamo e del Canto Alto	Versanti collinari	CB2
	Fondovalle	CB3
Val Brembana 2020	Alta montagna	VB1
	Media montagna	VB2
	Fondovalle	VB3
Valle Seriana e Laghi Bergamaschi	Sommitale	VSL1
	Versante	VSL2
	Fondovalle	VSL3

- L'IDENTIFICAZIONE DEI PRINCIPALI STAKEHOLDER È STATA AFFIDATA AI 4 GAL PARTNER MEDIANTE UN METODO EMPIRICO, IN VIRTÙ DELLA LORO CONOSCENZA E OPERATIVITÀ DIRETTA SUI TERRITORI INTERESSATI. IL METODO HA RIGUARDATO L'UTILIZZO DI UNA SCHEDA DI RILEVAZIONE AD HOC PROPOSTA DAL GRUPPO DI RICERCA E RIPORTANTE UNA SERIE DI INFORMAZIONI UTILI SIA A IDENTIFICARE IL SOGGETTO SIA A DEFINIRNE LA POSIZIONE RISPETTO AL SE

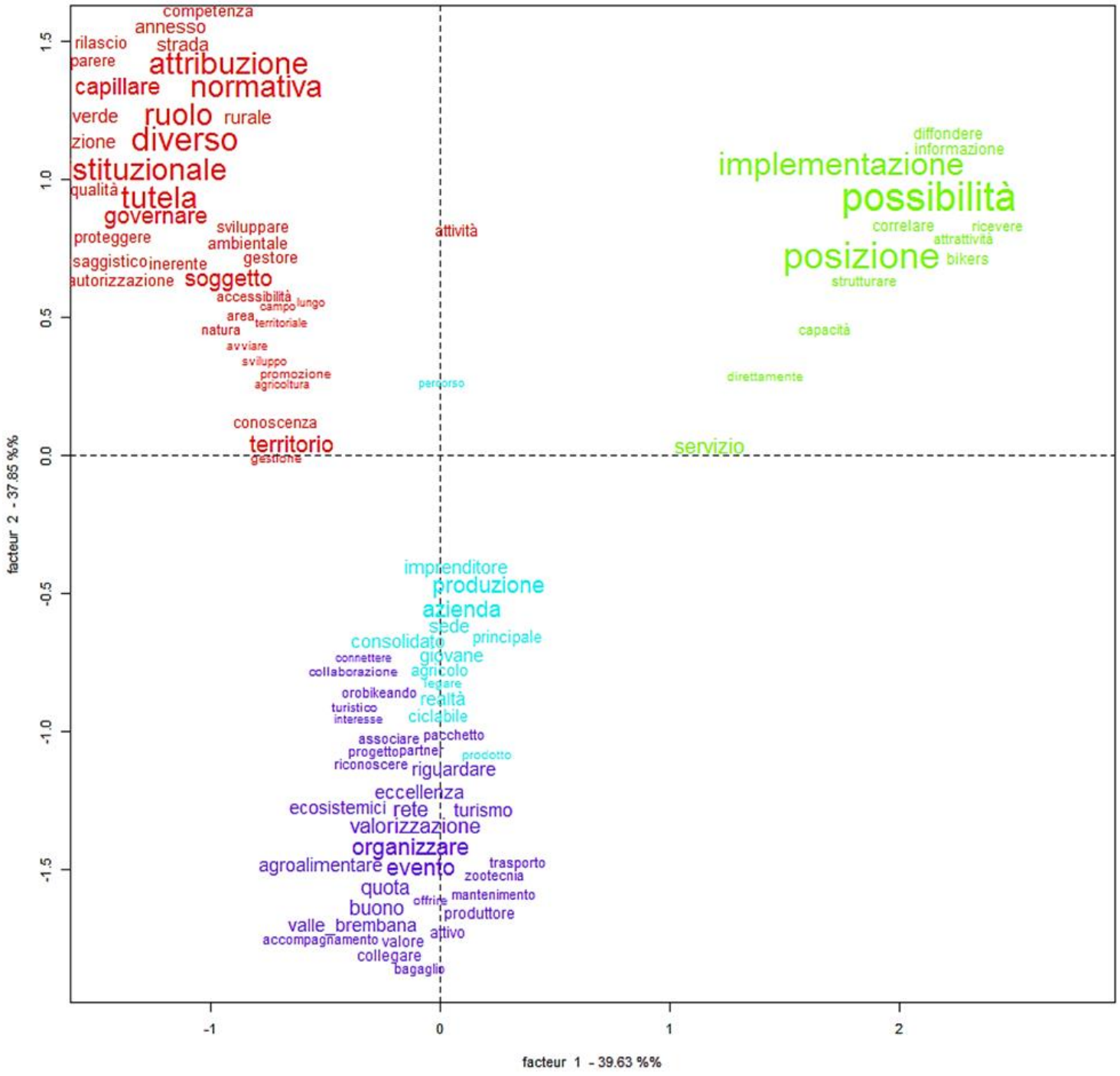
GAL	Unità Territoriale	Tipo STK	Cod_STK_Analisi T
Colli di Bergamo e del Canto Alto	CB2	Agriturismo	01
Val Seriana e Laghi Bergamaschi	CB3	Altro	02
Valle Brembana 2020	VB1	Associazione ambientalista	03
Valtellina - Valle dei Sapori	VB2	Associazione di categoria/Consorzio	04
	VB3	Attività ricettiva	05
	VSL1	Azienda agricola	06
	VSL2	Cooperazione	07
	VSL3	Ente locale (Comune, Parco, ecc..)	08
	VT1	Impresa agroalimentare	09
	VT2	Impresa di servizi	10
	VT3	Impresa ristorazione	11
		Servizi sportivi (società, negozi, ecc..)	12

CLASSI	
CA	Collaborazione Attesa
IA	Impegno Atteso
OP	Opportunità
PD	Punti di Debolezza
PF	Punti di Forza
RI	Rischi

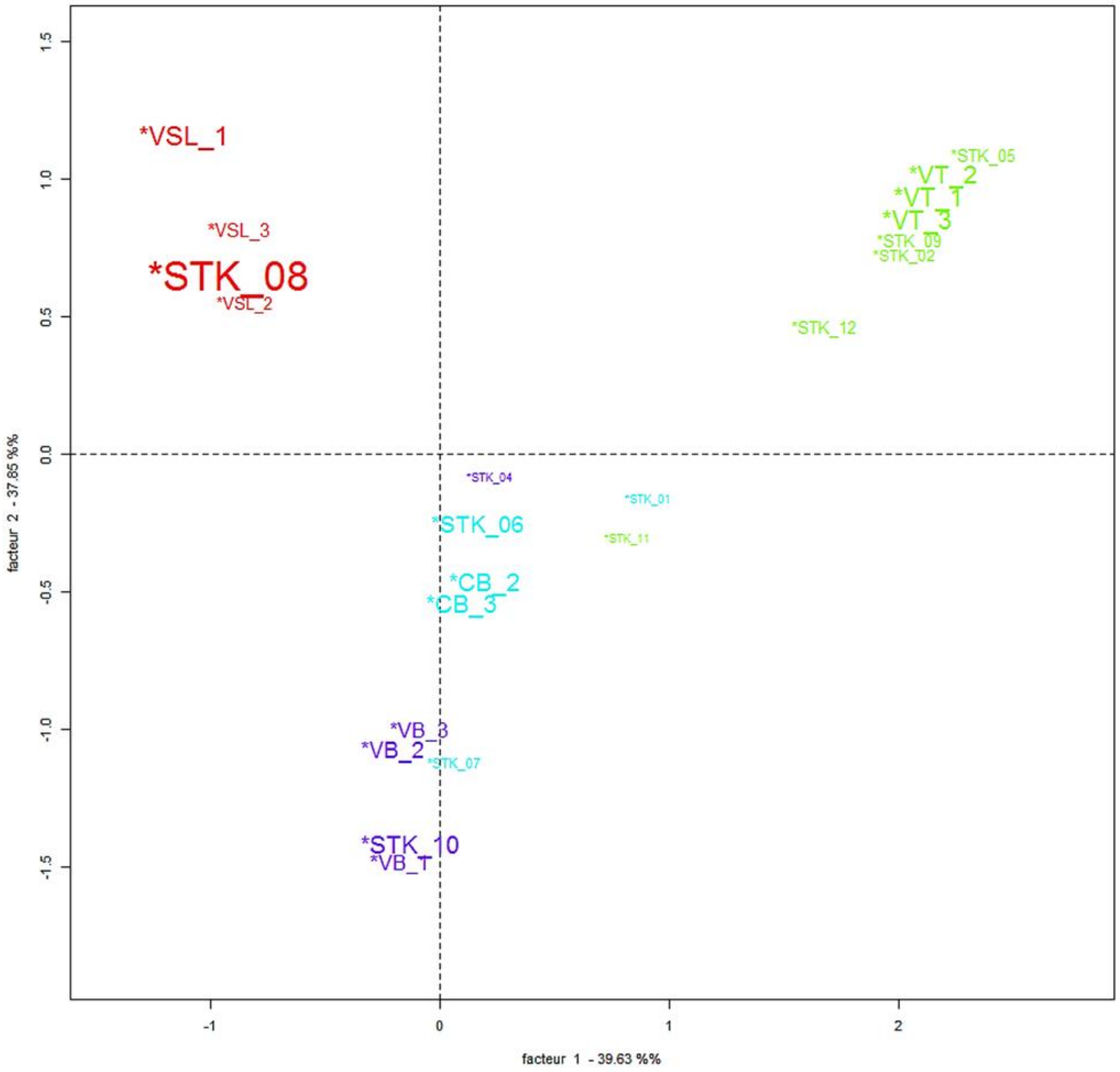
SWOT

268 schede:
11 unità territoriali X Tipo STK X Classi

Punti di Forza



Punti di Forza



SMART RURAL HUB

PF

Cluster 1

****** *STK_08 *VSL_3**

score : 418.63

capillare conoscenza di tutto il **territorio** e dei **soggetti** che lo **governano ruolo istituzionale** nella **tutela** del **territorio** attraverso le **diverse attribuzioni normative** inoltre ha **sviluppato** nel tempo **diverse attività** di **promozione** dell **accessibilità** del **territorio** in_particolare attraverso l implementazione del percorso **rurale** della **strada verde** e delle **dotazioni annesse**

****** *STK_08 *VSL_3**

score : 358.61

frequente propensione alla collaborazione con gli ambiti limitrofi per finalità condivise **ruolo istituzionale** nella **tutela** del **territorio** attraverso le **diverse attribuzioni normative** di **area protetta competenze** per **rilascio** di **autorizzazioni paesaggistiche gestore** di **aree natura** 2000 altri **pareri inerenti** la **qualità ambientale**

****** *STK_08 *VSL_2**

score : 358.61

ruolo istituzionale nella **tutela** del **territorio** attraverso le **diverse attribuzioni normative** di **area protetta competenze** per **rilascio** di **autorizzazioni paesaggistiche gestore** di **aree natura** 2000 altri **pareri inerenti** la **qualità ambientale**

Potenziali applicazioni in Agricoltura

- Schede Fabbisogni delle Postazioni Regionali - 95 schede fabbisogno prodotte nel complesso nel periodo da luglio 2017 a dicembre 2020 – Text Mining

IRaMuTeQ

R Interface for multidimensional analysis of texts and questionnaires

& python

http://www.iramuteq.org/

https://www.r-project.org/

Potenziali applicazioni: Analisi Schede Fabbisogni Postazioni Regionali

Prima di procedere con la loro mappatura, le schede fabbisogno sono state raggruppate e catalogate *ex ante* secondo una serie di criteri, alcuni oggettivi e altri maggiormente basati sull'interpretazione del gruppo di lavoro. I criteri ritenuti oggettivi sono:

- titolo della scheda;
- data di avvio della ricognizione;
- semestre di pubblicazione della scheda nel portale web della RRN;
- soggetto richiedente (AdG regionale, Ministero, ecc.);
- eventuale misura/operazione del PSR cui si riferisce il fabbisogno.

A tali criteri si aggiungono poi tre voci utili a identificare il contenuto specifico di ciascun fabbisogno; esse sono:

- tematica;
- parole chiave di I livello;
- parole chiave di II livello.

I due ordini di parole chiave sono gli elementi principali che servono a identificare i gruppi omogenei in cui raggruppare i fabbisogni, mentre la tematica è un campo descrittivo più articolato, utile nel caso si renda necessaria definire con maggior dettaglio i contenuti delle schede. Nel paragrafo successivo si riporta in maggior dettaglio il contenuto dei due livelli di parole chiave.

Potenziali applicazioni: Analisi Schede Fabbisogni Postazioni Regionali

Parole chiave di I livello

Il primo ordine di parole chiave ricalca indicativamente le diverse fasi di implementazione di un Programma di sviluppo rurale, dallo stadio iniziale di redazione/pianificazione degli interventi al momento finale di valutazione dei risultati. Il significato delle chiavi di I livello è illustrato in Tabella 1.

Tabella 1. Parole chiave di I livello

Chiave di I livello	Significato
Programmazione	Si riferisce alle fasi/processi di definizione e/o redazione di una parte del programma: ad es. i contenuti di una misura, il livello di un pagamento, un criterio di selezione o ammissibilità.
Attuazione	Si riferisce alla fase di esecuzione del programma, con particolare attinenza a ciò che genera un avanzamento del programma stesso, come ad es. la pubblicazione di un bando.
Gestione	Si riferisce anch'essa alla fase di esecuzione del programma, ma con focus sugli aspetti più amministrativi quali lo svolgimento di un'istruttoria, l'esecuzione dei controlli sui beneficiari di una misura, l'approvazione di una graduatoria.
Rendicontazione	Si riferisce alle problematiche connesse con il pagamento dei beneficiari o l'ammissibilità delle spese.
Monitoraggio/Valutazione	Si riferisce a problematiche e fabbisogni inerenti al calcolo degli indicatori, la redazione della RAA, ecc.
Comunicazione	Si riferisce a tutte le attività o elementi finalizzati a pubblicizzare il PSR, i suoi contenuti, i risultati.

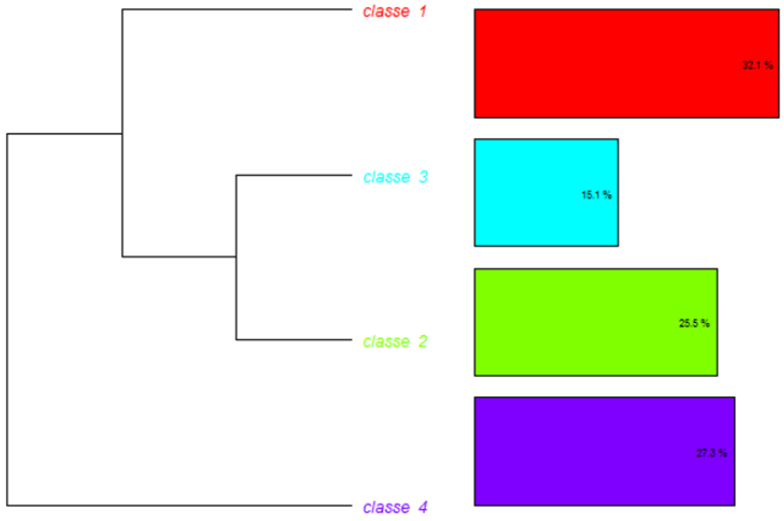
Parole chiave di II livello

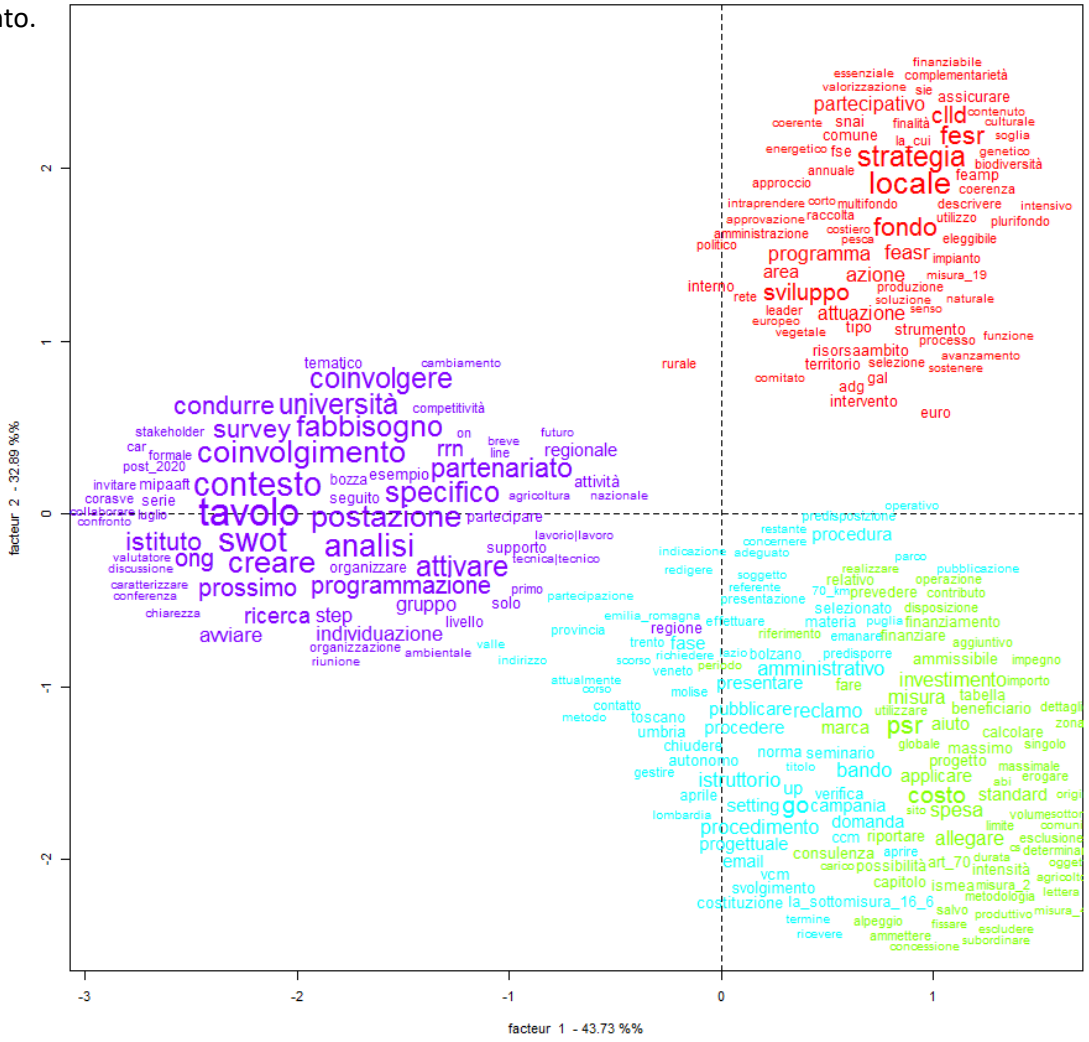
Come poc'anzi accennato, si tratta di termini che hanno la funzione di identificare più puntualmente il fabbisogno o la problematica affrontata. A titolo di esempio, sotto la chiave di I livello "**Attuazione**" si possono trovare più chiavi di II livello come *gruppi operativi*, *bandi*, *diversificazione*, ecc. Allo stesso modo, una parola chiave di II livello quale, ad esempio, "**banda larga**" può afferire a più lemmi di I livello, a seconda che il fabbisogno affrontato riguardi la fase programmatoria (in questo caso della Misura 7.5 del PSR) piuttosto che la gestione di un determinato aspetto amministrativo.

	Key - 1		Key - 2	Key-3
Parole Chiave I livello	Cod num	Parole Chiave II livello	Cod num	Anni
Attuazione	01	CdS	01	01=2017
Comunicazione	02	Controlli	02	02=2018_I
Gestione	03	Costi standard	03	03=2018_II
Monitoraggio/Valutazione	04	Credito	04	04=2019_I
Programmazione	05	De minimis	05	05=2019_II
Rendicontazione	06	Definizioni	06	
		Diversificazione	07	
		Durata informazioni	08	
		E-learning	09	
		Gruppi Operativi	10	
		Istruttorie	11	
		Multifondo	12	
		Pagamenti a ettaro	13	
		Pianificazione territoriale	14	
		Post 2020	15	
		Procedure	16	
		RAA	17	
		Referenti	18	
		Sovvenzione globale	19	
		Spese ammissibili	20	
		Spese personale	21	
		Spot radiofonici	22	
		Tabelle ettaro/coltura	23	
		Transizione	24	
		Trascinamenti 07-13	25	
		Zonizzazione / Definizioni	26	
		Performance framework	27	
		Piano di Comunicazione	28	
		Indicatori	29	
		CUP	30	
		Banda Larga	31	
		Bandi	32	
		Strumenti Finanziari	33	

Osservando più in dettaglio quali sono le parole che compongono i cluster, si possono identificare quattro temi polarizzanti:

1. rosso: strategia locale, approccio territoriale;
2. verde: implementazione, operatività;
3. azzurro: gestione, amministrazione;
4. viola: programmazione, logica di intervento.





Cagliero R., Chiozzotto F., Gabrieli G., Longhitano D., Rossi N., Vassallo M. (2021). **Fabbisogni delle Postazioni Regionali: Una rappresentazione dinamica**. CREA-PB Technical Report. ISBN 9788833851464.

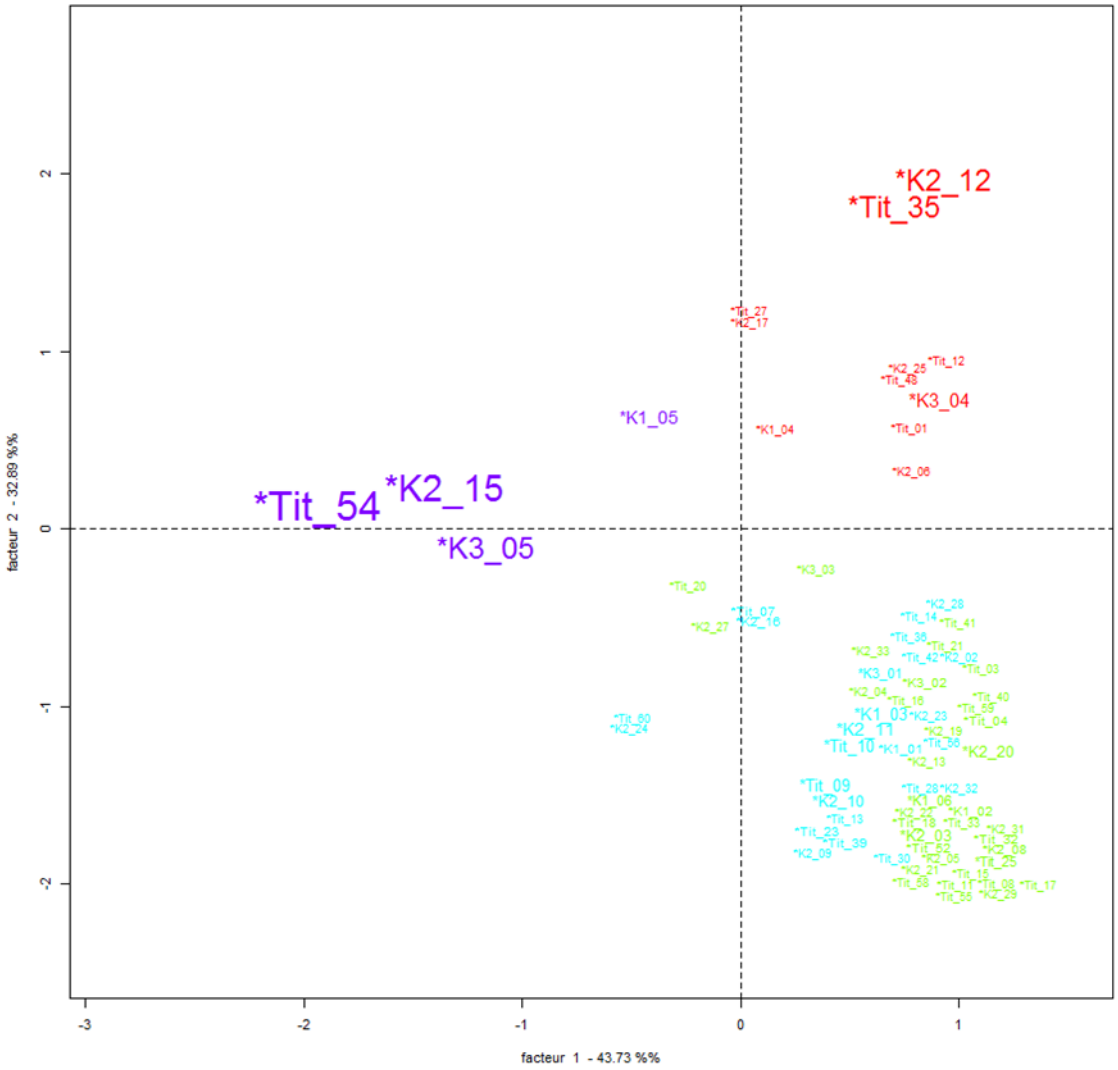
https://www.reterurale.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/23260

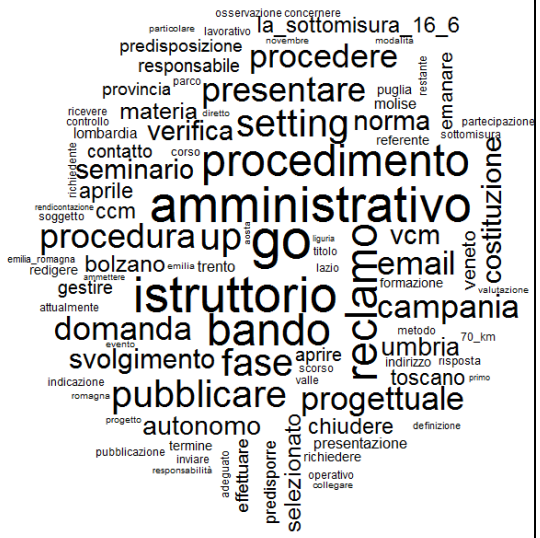
Come si può notare, il cluster 4 (viola) è identificato da pochissimi elementi dal peso particolarmente elevato, in particolare:

- Tit_54 – Nuova programmazione
- K2_15 – Post 2020 (chiave di II livello)
- K3_05 – II semestre 2019

Anche nel cluster 1 (rosso), pur con una predominanza meno netta, è possibile individuare una scheda fabbisogno dal peso più significativo rispetto alle altre:

- Tit_35 – CLLD Multifondo
- K2_12 – Multifondo



Cluster 1 Strategia locale, approccio territoriale	Cluster 2 Implementazione, operatività
K1: 05 = Programmazione **K2**: 12 = Multifondo; 06 = Definizioni; 25 = Trascinamenti 07-13 **K3**: 04 = 2019-I	**K1**: 06 = Rendicontazione **K2**: 20 = Spese ammissibili; 03 = Costi standard **K3**: 02 = 2018-I
Cluster 3 **Gestione, amministrazione**	**Cluster 4** **Programmazione, logica di intervento**
K1: 03 = Gestione **K2**: 11 = Istruttorie; 10 = Gruppi Operativi **K3**: 01 = 2017-II	**K1**: 05 = Programmazione **K2**: 15 = Post 2020 **K3**: 05= 2019-II
	

Potenziali applicazioni: Analisi Schede Fabbisogni Postazioni Regionali

Cluster 1 – Strategia locale, approccio territoriale	
***Tit_35 *K1_05 *K2_12 *K3_04**	il **sostegno** dei **fondi feasr** e **fesr** allo **sviluppo locale** di **tipo partecipativo clld** sarà **coerente** e coordinato tra i due **fondi** interessati tramite l'attuazione congiunta da parte delle due **adg** dello **strumento clld** attraverso il consolidamento di attività
Score: 508.29	
***Tit_35 *K1_05 *K2_12 *K3_04**	in ottemperanza all'art_**33** par_3 del **reg_ue_1303_2013** le **strategie** di **sviluppo locale** di **tipo partecipativo** saranno **selezionate** ed approvate da un **comitato istituito** a tale scopo dalle **adg** responsabili dei due **fondi fesr** e **feasr**
Score: 479.13	
***Tit_35 *K1_05 *K2_12 *K3_04**	scheda **misura_19** in considerazione della flessibilità dell'approccio **leader** i **gal** potranno prevedere **interventi** riconducibili ad altre misure **feasr** e ad altri **programmi** finanziati da **fondi sie** purché **coerenti** con gli **ambiti** tematici selezionati nelle **strategie** di **sviluppo locale**
Score: 470.08	

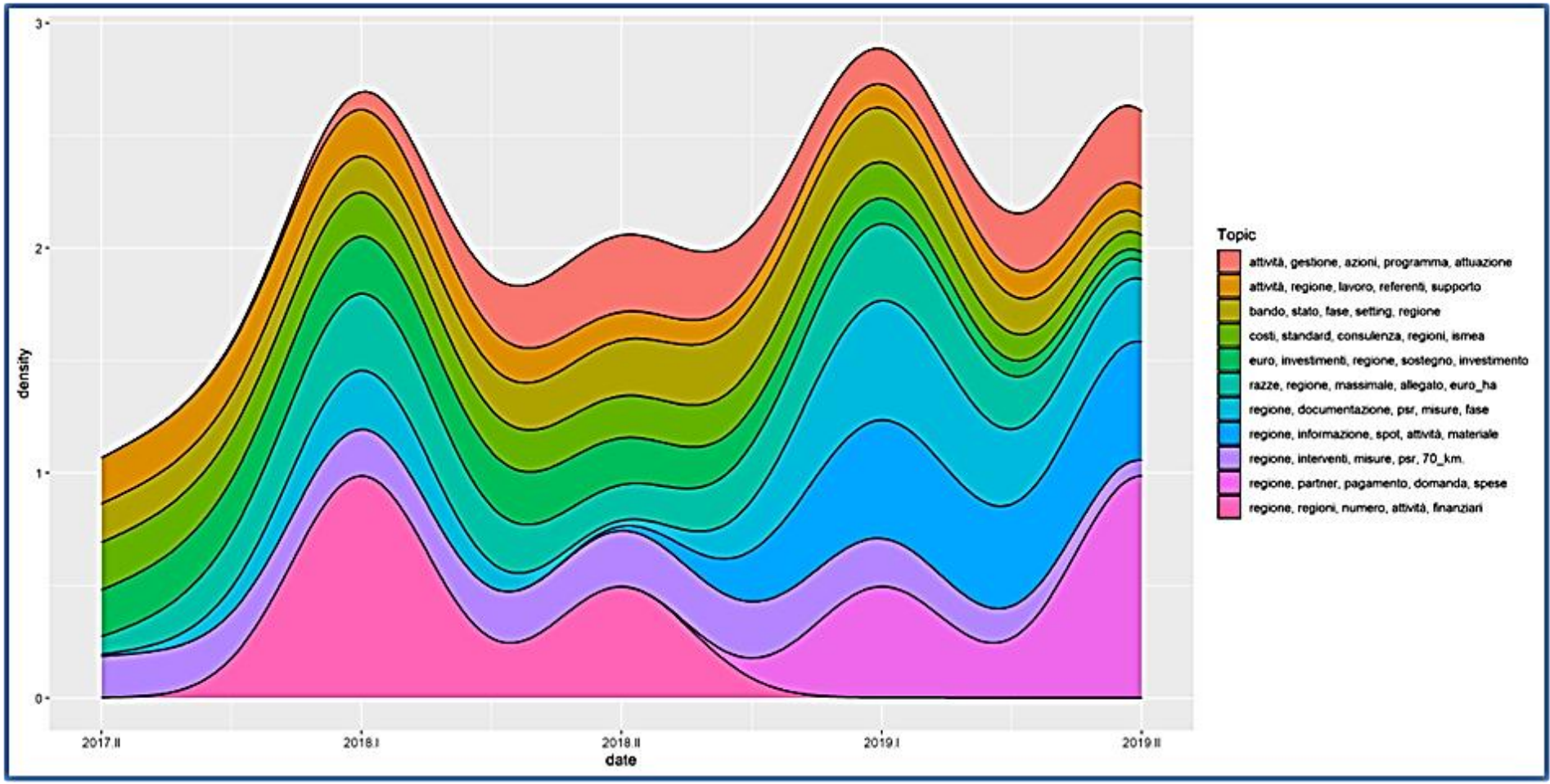
Cluster 2 – Implementazione, operatività	
***Tit_59 *K1_06 *K2_19 *K3_05**	in particolare, il **psr riporta** se il **progetto** contiene **investimenti produttivi** di cui all'allegato I del Trattato **finanziati** dalla misura_4 si **applicano** le **intensità massime** di **aiuto** della medesima **misura** maggiorata di un **ulteriore** 20_percento come da **allegato** ii del reg_1305_2013
Score: 339.53	
***Tit_59 *K1_06 *K2_19 *K3_05**	per quanto riguarda l'**importo globale** in **Lombardia** è **applicato** per la_sottomisura_16_1_01 sottomisura_16_2_01 e per la_sottomisura_16_4_01 nel **psr** si stabilisce che qualora i **costi** del **progetto facciano riferimento** ad un'altra **misura** sarà rispettato il **massimale** dell'**importo** e dell'**intensità** del sostegno da essa stabilito
Score: 295.42	
***Tit_03 *K1_05 *K2_20 *K3_01**	il **psr** non **prevede deroghe** per gli interventi fuori **regione** Piemonte il **psr** non **fa** alcun **riferimento** agli interventi fuori **regione** in particolare le **disposizioni** per le **misura_4_1** misura_4_2 misura_6_1 misura_6_4 non **ammettono finanziamenti** fuori **regione** così come gli **investimenti** strutturali in **alpeggio** e per le **misure** forestali
Score: 277.87	

Potenziali applicazioni: Analisi Schede Fabbisogni Postazioni Regionali

Cluster 3 – Gestione, amministrazione	
Tit_09 *K1_01 *K2_10 *K3_01	**selezionati** 22 gruppi per il **setting up Molise** il **bando** per il setting_up è **aperto Piemonte** il **Piemonte** è in **fase** di **valutazione** delle proposte **progettuali presentate** al **primo** step **Provincia** A**utonoma Bolzano**
Score: 301.25	
***Tit_09 *K1_01 *K2_10 *K3_01**	il secondo **bando** si è **chiuso** a metà **novembre** e le 6 **domande ricevute** sono ora in **fase** di **valutazione** Sardegna il 22 giugno 2017 è stato **pubblicato** il **bando** per la **fase** di **setting up** e si è **chiuso** il 15 **settembre scorso**
Score: 244.19	
***Tit_58 *K1_06 *K2_21 *K3_05**	anche la regione **Campania** adotta la medesima **procedura** della P**rovincia** di **Bolzano** per ogni **GO progetto** vi è un unico **soggetto** richiedente che **presenta** sia la **domanda** di aiuto sia quella di liquidazione a nome di tutto il gruppo
Score: 208.36	

Cluster 4 – Programmazione, logica di intervento	
Tit_54 *K1_05 *K2_15 *K3_05	e stato **attivato** uno **specifico gruppo** di lavoro come è **composto solo regione** con **università ONG istituti** Regione come si sta **coinvolgendo** il **partenariato** nessuna **riunione** qual è il **coinvolgimento ruolo** della **postazione** della **RRN** o della **postazione regionale** CREA
Score: 1002.43	
Tit_54 *K1_05 *K2_15 *K3_05	la **postazione** è stata **coinvolta** nell'organizzazione delle **prime attività** a **supporto** della **programmazione_2021_2027** Calabria la **regione** si è **attivata** verso la **prossima programmazione** su quali **step analisi** di **contesto swot individuazione fabbisogni**
Score: 883.63	
Tit_54 *K1_05 *K2_15 *K3_05	ha **contattato** la **postazione regionale crea** per un **supporto tecnico** scientifico all'elaborazione della **swot regionale** quali **attività** sono state **condotte** o si **condurranno esempio survey tavoli ricerche specifiche analisi** delle **bozze** di **regolamento** e relativi **documenti** di lavoro
Score: 863.78	

Potenziali applicazioni: Analisi Schede Fabbisogni Postazioni Regionali – **ACCENNI al TOPIC MODELING**



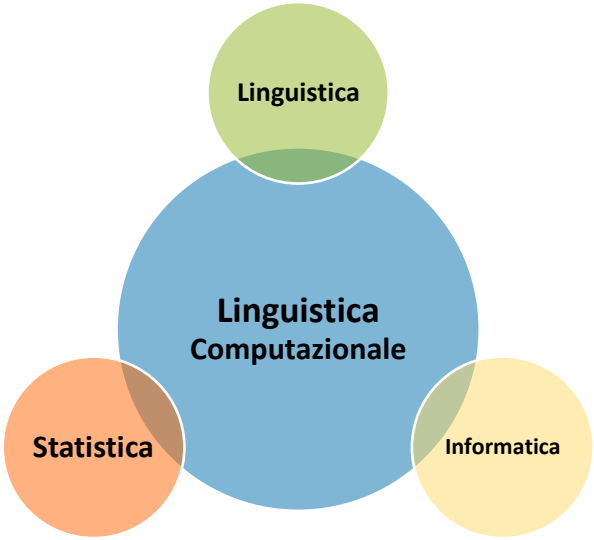
OPINION MINING:

SENTIMENT ANALYSIS

Cos'è la Sentiment Analysis?

- **L'analisi del sentimento, conosciuta come *sentiment analysis o opinion mining*, viene definita come quella applicazione della linguistica computazionale "…che si occupa di classificare ed analizzare le opinioni soggettive e le emozioni espresse in un frammento di linguaggio naturale" (Basile, 2020*; p.160). (es. misurare l'interesse di clienti verso un prodotto/servizio – ambito commerciale; conoscere le opinioni di utenti di social media – ambito giornalistico comunicativo)**

- **In termini più semplici: estrazione automatica del sentimento (o polarità) positivo, negativo, neutrale in un messaggio in linguaggio naturale.**



*Basile V. (2020). I Computer e il Linguaggio Naturale. Ithaca: Viaggio nella Scienza, 16, 151-165. http://siba-ese.unisalento.it/index.php/ithaca/article/view/23009

Cos'è la Sentiment Analysis? Come si fa?

- Modelli non supervisionati basati sul dizionario (lessici affettivi: elenco di parole (forme o lemmi) associate ad uno o più punteggi numerici o categoriali.):
  - *MAL (Morphologically-inflected Affective Lexicon) and **Weighted-MAL -> DI.UniTo & CREA-PB
- Modelli supervisionati di apprendimento automatico da testi annotati – machine learning
  - Reti neurali: Google BERT (Bidirectional Encoder Representations from Transformers)
- Modelli ibridi
  - BERT+WMAL: A Hybrid Lexicon-based and Neural Approach for eXplainable Polarity Detection (submitted to Information Processing & Management – Elsevier IF:6,222) - >UniBA, UniTO, CREA-PB.

*Vassallo M., Gabrieli G., Basile V., Bosco C. (2019). The tenuousness of lemmatization in lexicon-based sentiment analysis. Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019). http://ceur-ws.org/Vol-2481/

**Vassallo M., Gabrieli G., Basile V., Bosco C. (2020). Polarity imbalance in lexicon-based sentiment analysis. Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020). http://ceur-ws.org/Vol-2769/

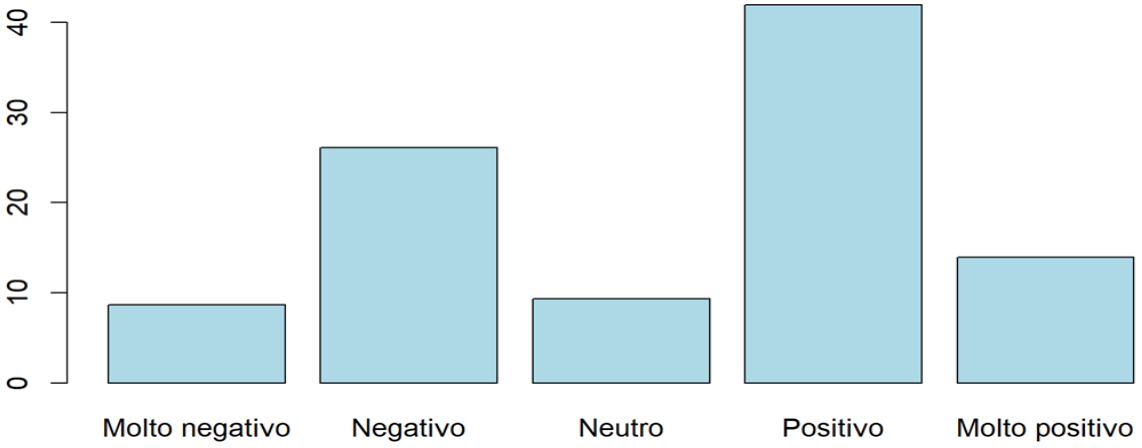
Potenziali applicazioni: CREAgritrend – SENTIMENT ANALYSIS

PARTI SOCIALI
- Cia_Agricoltura
- coldiretti
- Confagricoltura
- UciAgricoltura
- assorurale
- AIABfederale
- Agrinsieme

MEDIA
- AgriculturaIT
- Agricolae1
- agronotizie
- AGRAPRESS
- AgricolturaNews

© Twitter

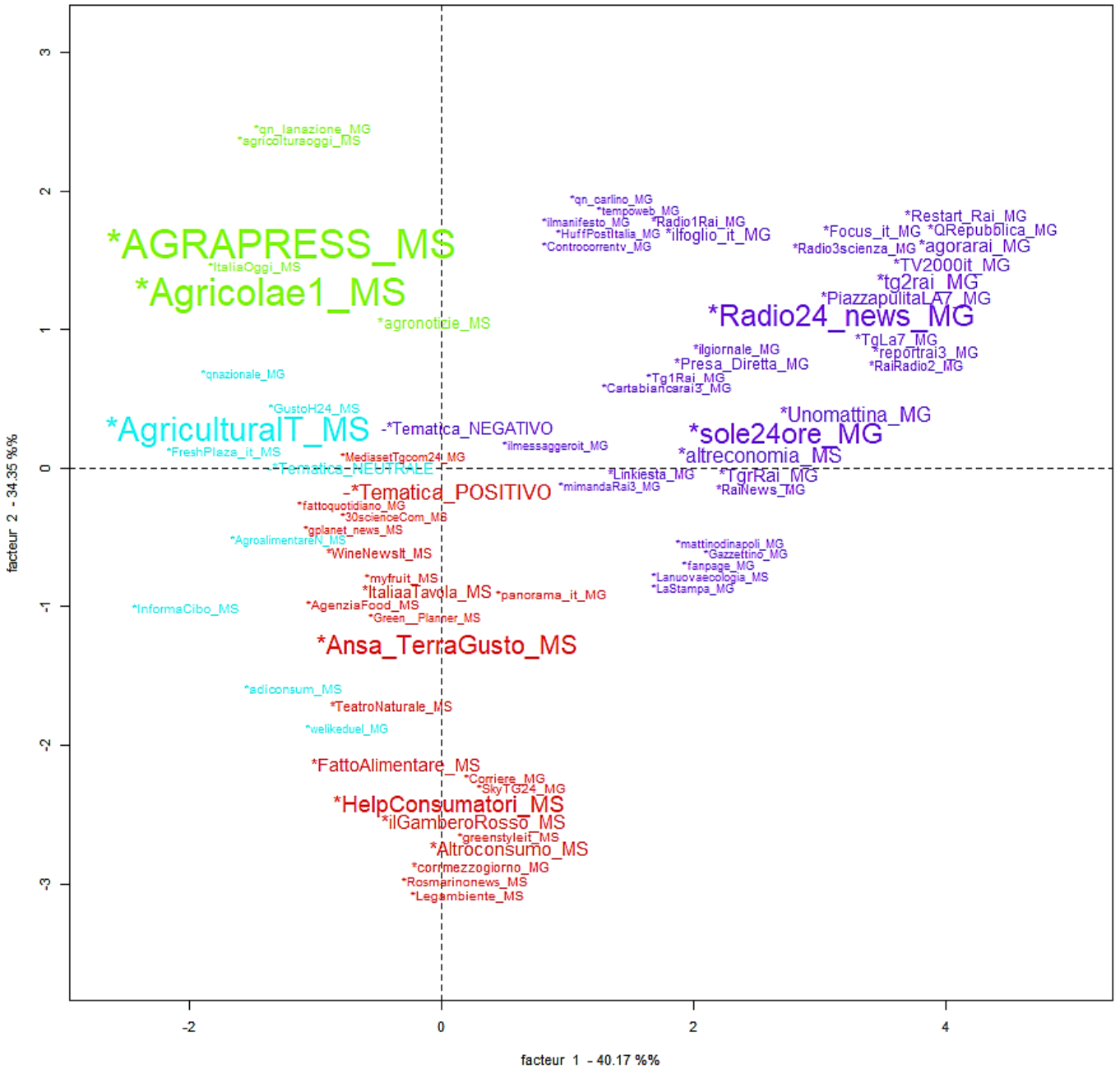
3.950 tweets, raccolti tra il 01 luglio e il 15 settembre 2019



Potenziali applicazioni nell'Agroalimentare: Text & Opinion Mining

© Twitter

- **MG=MEDIA_GENERALISTA**
- **MS= MEDIA_SPECIALIZZATO**



Potenziali applicazioni: **RICA-SENTIMENT**

- **Descrizione:** indagine sperimentale, a costo zero, volta a identificare e quantificare le opinioni degli imprenditori agricoli italiani che parteciperanno all'indagine campionaria **RICA*** 2019 (11.100 imprenditori) sulle politiche agricole comunitarie nel loro diretto contesto territoriale e di comparto.

*Rete di Informazione Contabile Agricola
https://rica.crea.gov.it/index.php

- **Obiettivi:**
  - Identificare e quantificare in termini di orientamento positivo, negativo, neutrale le opinioni degli imprenditori agricoli RICA sulle politiche agricole comunitarie nel loro diretto contesto territoriale e di comparto.
  - Individuare diverse prospettive, nuovi punti di vista e valutazioni sulle politiche agricole comunitarie espresse direttamente dagli imprenditori agricoli RICA.
- **Modalità:** invio tramite PEC di due/tre domande sulla PAC pre e post 2020 senza limiti di testo nelle risposte
- **Analisi dei dati:** Text & Opinion Mining

1) Quali vantaggi hanno prodotto per la sua azienda gli aiuti della PAC 2014-2020?

2) Quali novità hanno portato maggior beneficio (es. greening, gestione del rischio, aiuti ai giovani)?

3) E quali sono invece le criticità che ha potuto riscontrare?

4) Alla luce della nuova riforma PAC post 2020 cosa si aspetterebbe in particolare per la sua azienda dalla nuova programmazione?

5) Alla luce dell'emergenza covid 19 cosa si aspetterebbe in particolare per la sua azienda dalla nuova programmazione?

Attività di Ricerca:

- Social Media Analytics in Agriculture – SMs-Agri - https://smsagri.crea.gov.it/

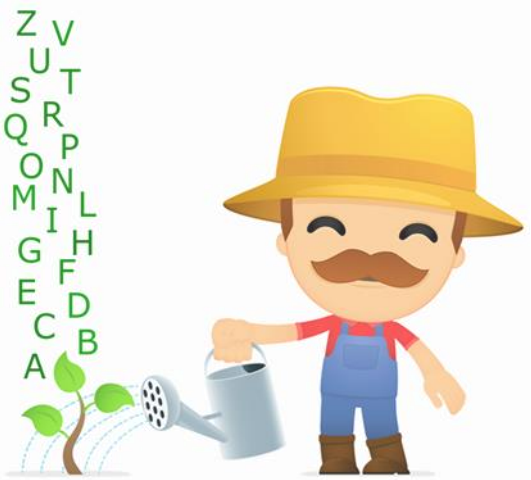
Prospettive Future:

- Accordo di collaborazione (art.15) con il Dipartimento di Informatica dell'Università degli Studi di Torino – DI.UNITO: https://informatica.unito.it/do/home.pl
- Gruppo di ricerca Computazione Centrata sui Contenuti (Content-Centered Computing group, https://cs.unito.it/do/gruppi.pl/Show?_id=453y

di.unito.it

Content Centered Computing

AILC Associazione Italiana di Linguistica Computazionale

http://www.ai-lc.it/

Grazie per l'attenzione!