

ENUMERA

STORIE DI DOCUMENTI, DI LIBRI E DI PAROLE

3

Direttore

Roberto GUARASCI
Università della Calabria

Comitato scientifico

Mauro GUERRINI
Università degli Studi di Firenze

Giovanni SOLIMINE
Università di Roma La Sapienza

Laurence FAVIER
Université Charles de Gaulle Lille 3

Madjid IHADJADENE
Université Paris 8

ENUMERA

STORIE DI DOCUMENTI, DI LIBRI E DI PAROLE

Collana di scienze del testo e del documento aperta alla contaminazione di tutte quelle discipline che, a vario titolo, studiano il potere della parola e dell'informazione bibliografica e documentale e la sua capacità di perdurare nel tempo.

Erika Pasceri

**Strutture tassonomiche e linguaggi
specialistici in ambito Biomedico**

Prefazione di
Mauro Guerrini



Copyright © MMXIV
ARACNE editrice S.r.l.

www.aracneeditrice.it
info@aracneeditrice.it

via Raffaele Garofalo, 133/ A-B
00173 Roma
(06) 93781065

ISBN 978-88-548-7175-5

*I diritti di traduzione, di memorizzazione elettronica,
di riproduzione e di adattamento anche parziale,
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

*Non sono assolutamente consentite le fotocopie
senza il permesso scritto dell'Editore.*

I edizione: giugno 2014

Indice

- 9 *Prefazione*
di Mauro Guerrini
- II *Introduzione*
- 15 **Capitolo I**
Il contesto internazionale e nazionale
- 1.1. Le malattie rare: Il contesto internazionale di riferimento, 15 –
1.1.1. *Le associazioni sulle malattie rare a livello internazionale*, 18 –
1.1.2. *I principali studi a livello internazionale sulle malattie rare*, 27 –
1.2. Analisi del contesto nazionale, 28.
- 31 **Capitolo II**
L'importanza degli standard nell'informatica medica
- 2.1. L'Unified Medical Language System, 31 – 2.1.1. *Il Metathesaurus*, 32 – 2.1.2. *Semantic Network*, 33 – 2.1.3. *Specialist Lexicon*, 34 –
2.2. L'analisi terminologica, 35 – 2.2.1. *Acquisizione e analisi dei dati provenienti da Orphanet, dall'ORDR e dalla NORD*, 36 –
2.2.2. *Le terminologie target: OMIM, SNOMED CT, MeSH, ICD10*, 38 –
2.2.3. *La metodologia di lavoro*, 50 – 2.2.4. *Risultati della mappatura verso l'UMLS*, 56.
- 61 **Capitolo III**
Il modello di integrazione del Registro Nazionale delle Malattie Rare
- 3.1. Avvio della costruzione del thesaurus: il Registro Nazionale delle Malattie Rare, 61 – 3.2. Dal Registro al thesaurus, 66 –
3.2.1. *Il thesaurus e le funzioni d'uso*, 69 – 3.2.2. *Lo strumento utilizzato*, 73.

75 *Conclusioni*

83 *Appendice*

93 *Bibliografia*

Prefazione

di MAURO GUERRINI

Il dominio della sanità e delle scienze della vita è tra i più complessi e articolati dal punto di vista documentario, informativo e terminologico; è, pertanto opportuno, anzi necessario, poter disporre di adeguati strumenti di strutturazione della conoscenza registrata, così da poterla padroneggiare e rappresentare. Alla complessità del tema si aggiungono la ricchezza lessicale insita in ciascuna lingua e l'indispensabile confronto con le caratteristiche specifiche della globalità delle lingue, che certamente aggrava la situazione e rende complicato ogni lavoro di indicizzazione. Per ottenere un risultato soddisfacente, occorrono tre prerogative essenziali: acquisire grande competenza scientifica, gestire in modo coordinato le iniziative internazionali e disporre di tecnologie digitali aggiornate e funzionali.

In tempi recenti la comunità scientifica ha finalmente riconosciuto l'importanza delle terminologie e la necessità della loro standardizzazione, in un contesto che ne prevedesse l'interoperabilità; sono state, perciò, avviate politiche condivise di gestione dell'informazione di dominio che prevedano il fondamentale dialogo tra i progetti realizzati nei singoli contesti nazionali. Non sempre, tuttavia, il coordinamento riesce a coprire tutte le vastissime sfaccettature del dominio, anche per l'ovvia limitatezza delle risorse umane e finanziarie disponibili.

Erika Pasceri analizza un caso di nicchia rispetto l'ampio campo delle specializzazioni mediche esistenti; rappresenta la strutturazione della conoscenza delle patologie rare, un settore poco esplorato. Proponendone una terminologia controllata. Nell'ambito delle scienze dell'informazione e, in questo caso

anche in ambito medico, il modello proposto rappresenta un valido strumento per veicolare la standardizzazione di informazioni sia in fase di input sia per l'accesso guidato da parte degli utenti. Il web 2.0 ha ridimensionato il sistema informativo tradizionale in ogni campo del sapere; la rete mette a disposizione innumerevoli sistemi e modalità per il reperimento delle informazioni. Spesso, tuttavia, il controllo qualitativo è carente e ciò inficia il recupero delle informazione, ottenendo notizie errate e incomplete o, al contrario, un eccesso di rumore e l'impossibilità a controllare la veridicità dell'asserzione. La possibilità di rappresentare le strutture tassonomiche in linguaggi formali apre la strada a nuovi e possibili utilizzi delle risorse terminologiche e thesaurali nell'ambito degli artefatti ontologici a servizio della crescente comunità globale degli utenti. In questo senso e in questa prospettiva ogni nuova realizzazione di strutture controllate di classificazione è sicuramente meritevole di apprezzamento e di plauso.

Introduzione

Il lavoro rappresenta il frutto dell'attività di ricerca condotto durante il XXV ciclo di dottorato in *Scienze bibliografiche, archivistiche e documentarie e per la conservazione e restauro dei beni librari ed archivistici*, svoltosi presso l'Università di Udine nel triennio 2010-2012.

L'obiettivo principale è stato quello di analizzare la rappresentazione terminologica delle malattie rare e di costruire un modello di classificazione per tali patologie a supporto non solo per i professionisti del settore, ma anche e soprattutto per gli utenti/assistiti affetti da sindromi rare affinché gli stessi possano utilizzarlo come strumento per la ricerca di informazioni cliniche utili ai propri percorsi di diagnosi e cura.

Le persone affette da malattie rare si trovano troppo spesso ad affrontare problemi come l'impossibilità di ottenere in tempi rapidi una corretta diagnosi, la mancanza di informazioni esauritive sulla patologia e, spesso, la mancanza di casistica specifica. Tutto ciò non sempre e non solo per oggettiva mancanza di informazioni ma, spesso, per difetto o inesistenza di strumenti di accesso all'informazione stessa. Prendendo come esempio, la sindrome WAGR,¹ disturbo patologico estremamente raro legato alla delezione del cromosoma 11 (11q) e caratterizzata dal tumore di Wilms, da anomalie genito-urinarie, e da ritardo nello sviluppo, si scopre che diversi sono stati i termini conosciuti per identificarla, tra cui, appunto, l'acronimo WAGR; termini che si riferiscono alla sua origine genetica (sindrome da delezione 11q, Monosomia 11p13) o alla combinazione di eventi che ne conseguono (tumore di Wilms, anomalie genitourinarie,

1. <http://www.wagr.org/>.

sindrome da ritardo mentale) senza che siano sempre esplicitate le relazioni di sinonimia esistenti tra i termini.

D'altro canto il problema della non univocità semantica è una questione di non poca rilevanza nel momento in cui si deve compilare, ad esempio, la scheda di dimissione ospedaliera (SDO) e ci si trova davanti il problema della non esistenza di un termine non ambiguo per identificare una determinata patologia e, ci si vede, quindi, costretti ad usare termini generici, sintomatologie o perifrasi².

I dati raccolti attraverso le SDO costituiscono un prezioso strumento di conoscenza, di valutazione e di programmazione delle attività di prevenzione, diagnosi e cura ma rischiano di essere fortemente limitati se non spesso falsati dalla poca accuratezza descrittiva derivante da un linguaggio non standardizzato.

La necessità di gestire l'informazione e di creare sistemi di accesso e di organizzazione della conoscenza nel settore specifico trova la sua ragion d'essere nella particolare situazione in cui versa il nostro Paese, caratterizzata dall'assenza o dalla poca accuratezza di schemi di classificazione ai quali sia possibile fare riferimento per la compilazione delle SDO, per lo scambio non ambiguo di informazione, per l'interscambio dei dati in ambiente digitale, ecc.

L'obiettivo principale che ci si è proposti di raggiungere, propriamente terminologico, è, quindi, quello di indirizzare gli specialisti del dominio verso un utilizzo condiviso della terminologia specialistica, che possa garantire l'individuazione chiara ed univoca dei concetti corrispondenti, senza rischi di ambiguità. Partendo dal presupposto che esistono numerose

2. Per la codifica delle informazioni cliniche contenute nelle SDO in Italia si utilizza l'International Classification of Diseases 9th revision (ICD9-CM) ai sensi del D.M. del 26 luglio 1993. La codifica delle informazioni cliniche riportate nella SDO deve essere effettuata da personale sanitario (medico o infermieristico) adeguatamente formato. Spesso accade, tuttavia, che l'ICD9-CM non copra in modo esaustivo tutte le patologie nella loro complessità e, quindi, il compilatore della SDO si trova costretto ad utilizzare il codice più generico.

classificazioni dell'ambito biomedico, ed in particolare in quello delle malattie rare, il presente lavoro ha come obiettivo quello di costruire un modello che possa essere in un certo qual modo "unificatore" delle precedenti basi costruite.

La base terminologica di partenza è stato il Registro Nazionale delle Malattie Rare (RNMR) il riferimento utilizzato, ad oggi, nel contesto nazionale italiano, soprattutto dai professionisti del settore per la refertazione delle patologie rare. Il RNMR durante la fase di ricognizione e analisi dello stato dell'arte è risultato essere in alcuni casi poco preciso. È stato effettuato, quindi, il controllo semantico avvalendosi di specifici modelli di riferimento dei quali si daranno ulteriori dettagli nel secondo capitolo.

Per meglio comprendere le ragioni di questa scelta metodologica, si riporta un esempio estratto dal RNMR ed in particolare dalla categoria "*Malattie delle ghiandole endocrine, della nutrizione, del metabolismo e disturbi immunitari*", in cui compare il termine relativo alla *Sindrome di George*, inclusa tra le "*Immunodeficienze primarie*". In realtà la *Sindrome di George* appartiene alla categoria delle malattie cromosomiche, come si evince dalla sua definizione:

La diagnosi di sindrome di George è clinica. La sindrome di George è caratterizzata dall'insieme di diverse malformazioni: ipoplasia del timo e delle ghiandole paratiroidi, cardiopatia congenita di tipo troncoconale e dimorfismi del viso minori ma caratteristici. La sindrome velo-cardio-facciale si distingue per l'associazione di alterazioni cardiache di tipo troncoconale, palatoschisi o insufficienza del velo palatino, dimorfismi del viso e disturbi dell'apprendimento. È noto che queste due sindromi rappresentano espressioni cliniche della stessa condizione, in diverse età. La loro definizione si sovrappone a quella di microdelezione 22q11, che è presente in oltre il 95% dei pazienti. L'acronimo CATCH 22 è stato proposto per descrivere le caratteristiche cliniche della microdelezione 22q11 (Cardiac-Abnormal face-Thymus-Cleft palate-Hypocalcemia). L'evoluzione della sindrome è largamente condizionata dalla natura delle malformazioni congenite. L'ipocalcemia, spesso osservata nel periodo prenatale, di solito scompare, ma in alcuni bambini può per-

sistere un ipoparatiroidismo che deve essere trattato. L'insufficienza velofaringea spesso esita in voce nasale, anche senza palatoschisi, e può essere accompagnata a difficoltà nel linguaggio. Nella popolazione generale l'incidenza stimata della microdelezione 22q11 è circa 1/5000 nati. La microdelezione 22q11 si trasmette con modalità autosomica dominante e nel 10-20% dei casi uno dei due genitori è portatore della microdelezione. Tuttavia, nella maggior parte dei casi, la mutazione cromosomica è *de novo*³

Inoltre, sempre facendo riferimento al RNMR nessun sinonimo è stato associato a tale patologia, mentre ne esistono diversi⁴:

USE Sindrome di George
 UF CATCH 22
 Microdelezione 22q11
 Sequenza di George
 Sindrome da anomalie facciali e troncoconali
 Sindrome di Sedlackova
 Sindrome di Shprintzen
 Sindrome velo-cardio-facciale

L'esempio qui riportato evidenzia come con una struttura thesaurale ben definita e completa, che organizzi il complesso terminologico delle risorse disponibili, possa indirizzare l'utente, attraverso il reticolo di relazioni semantiche a raggiungere agevolmente l'informazione ricercata oltre che a facilitarne il mantenimento e l'aggiornamento.

3. [http://www.orphana.net/consor/cgi-bin/Disease_Search.php?lng=IT&data_id=126&Disease_Disease_Search_diseaseGroup=sindrome-di-George&Disease_Disease_Search_diseaseType=Pat&Malattia\(e\)/%20gruppo%20di%20malattie=Monosomia-22q11--Sindrome-di-DiGeorge-&title=Monosomia-22q11--Sindrome-di-DiGeorge-&search=Disease_Search_Simple](http://www.orphana.net/consor/cgi-bin/Disease_Search.php?lng=IT&data_id=126&Disease_Disease_Search_diseaseGroup=sindrome-di-George&Disease_Disease_Search_diseaseType=Pat&Malattia(e)/%20gruppo%20di%20malattie=Monosomia-22q11--Sindrome-di-DiGeorge-&title=Monosomia-22q11--Sindrome-di-DiGeorge-&search=Disease_Search_Simple).

4. La relazione thesaurale qui presentata per la "Sindrome di George" è stata da me costruita sulla base dello studio condotto e ampiamente descritto nel capitolo 2, par. 2.2.