Un modello a livelli per le edit su documenti testuali

*Versione 1.1*

*7 novembre 2018*

*Fabio Vitali*

# 1 Formato a tre livelli

1. Livello meccanico: due operazioni, INS e DEL, **rigorosamente** su testo (il markup viene considerato testo).
2. Livello strutturale: sei operazioni sul testo, 8 o 9 sulla struttura.
3. Livello semantico: svariate operazioni da decidere e considerare

Negli esempi seguenti, il testo in rosso rappresenta campi facoltativi (in generale non metterei by e timestamp nelle operazioni meccaniche).

# 2 Operazioni meccaniche

Ci sono solo due operazioni: INS e DEL. teniamo il nome così a tre lettere per distinguerlo dalle operazioni più significative a livello più alto.

le operazioni meccaniche agiscono ESCLUSIVAMENTE a livello della stringa di testo: operazioni sul markup sono comunque espresse come operazioni di INS e DEL di stringhe che contengono il markup. Di solito non hanno, ma opzionalmente possono avere indicazione di autore e timestamp. Gli id sono rigorosamente sequenziali, globali (indipendenti dalla versione a cui appartengono - iniziano da zero con la prima editazione e proseguono sequenzialmente) e usano una numerazione indipendente dagli altri livelli.

La posizione dell'inserimento viene contata sulla stringa dei caratteri del testo, considerando anche il markup e contandone ogni carattere. Il whitespace viene sempre considerato. INS e DEL sono sufficienti garantire la correttezza e la bidirezionalità dell'algoritmo di ricostruzione del testo. Operational Transformation ***non opera*** a questo livello, ma solo al livello successivo.

### 2.1 Inserimenti (INS)

|  |
| --- |
| **INS** |
| *Inserimento di una stringa di caratteri (incluso markup)* |
| {  "id": "edit-00001",  "op": "INS",  "pos": 80,  "content": "nuovo"  } |
| {  "id": "edit-00003",  "op": "INS",  "pos": 100,  "content": "</p><p>"  } |
| {  "id": "edit-00005",  "op": "INS",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "pos": 120,  "content": "new content"  } |

### 2.2 Cancellazioni (DEL)

|  |
| --- |
| **DEL** |
| *Eliminazione di una stringa di caratteri (incluso markup)* |
| {  "id": "edit-00002",  "op": "DEL",  "pos": 90,  "content": "vecchio"  } |
| {  "id": "edit-00004",  "op": "DEL",  "pos": 110,  "content": "part of the first paragraph.<p>Part of "  } |
| {  "id": "edit-00006",  "op": "DEL",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",  "pos": 130,  "content": "old content"  } |

# 3 Operazioni strutturali

Sono operazioni a livello strutturale sequenze di uno o più operazioni meccaniche che hanno un senso proprio sul documento anche se vengono svolte in più fasi o in locazioni diverse. Le operazioni strutturali possono risolversi all'interno di un unico nodo di testo (e parleremo allora di operazioni sul testo) o su strutture di markup complesse (e parleremo allora di operazioni sulla struttura. Le operazioni strutturali vengono svolte atomicamente (ad esempio, un wrap avviene atomicamente anche se corrisponde a due insert indipendenti). Questo significa che timestamp e autore delle operazioni strutturali sono indispensabili e si applicano a tutte le operazioni meccaniche sottostanti senza bisogno di specificarle. Operational Transformation agisce a questo livello.

Si noti che la stessa operazioni strutturale può avere due o più sequenze meccaniche diverse con lo stesso effetto. Ad esempio, l'operazione di WORDREPLACE "word" -> "words" può essere vista come l'INS di una 's', oppure come il DEL di "word" e l'INS di "words", ma dal punto di vista strutturale sono equivalenti.

## 3.1 Operazioni sul testo

Sono operazioni sul testo quelle che avvengono all'interno di uno stesso nodo di markup. Se un'operazione impatta uno o più nodi di markup, anche se assomiglia ad una operazione sul testo, va comunque considerata come un'operazione sulla struttura. Le operazioni di testo sono all'interno di una parola (ma solo se ne correggono lo spelling), operazioni su una parola, ed operazioni su più parole.

### 3.1.1 Punteggiatura (PUNCTUATION)

Si tratta di modifiche della sola punteggiatura senza incidere in alcuna maniera nel testo vero e proprio. Sono incluse nella modifica della punteggiatura anche quelle operazioni complesse in cui la parola immediatamente seguente al punto diventa maiuscola come effetto dell'interruzione della frase:

|  |  |
| --- | --- |
| **PUNCTUATION** | |
| *Modifica della sola punteggiatura.* | |
| {   "id": "structural-00010",  "op": "PUNCTUATION",  "old": "done; therefore",  "new": "done. Therefore",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00048",  "op":"DEL",   "pos": 123,   "content": "; t"   },{   "id" : "edit-00049",  "op":"INS",   "pos": 123,   "content": ". T"   }] } | |

### 3.1.2 Cambio dentro a parola (WORDCHANGE)

|  |  |
| --- | --- |
| **WORDCHANGE** | |
| *Modifica di una singola parola senza cambiamento della parola complessivo (cambio o correzione di uno spelling)*  *Tutti gli insert e i replace sono all'interno di una parola sola.* | |
| {   "id": "structural-00011",  "op": "WORDCHANGE",  "old": "giuoco",  "new": "gioco",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00050",  "op":"DEL",   "pos": 123,   "content": "u"   }] } | |
| {   "id": "structural-00011",  "op": "WORDCHANGE",  "old": "giuoco",  "new": "gioco",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00050",  "op":"DEL",   "pos": 121,   "content": "giuoco"   },{   "id" : "edit-00052",  "op":"INS",   "pos": 121,   "content": "gioco"   }] } | |
| {   "id": "structural-00012",  "op": "WORDCHANGE",  "old": "anlyze",  "new": "analyse",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00051",  "op":"INS",   "pos": 161,   "content": "a"   },{   "id" : "edit-00052",  "op":"DEL",   "pos": 164,   "content": "z"   },{   "id" : "edit-00053",  "op":"INS",   "pos": 164,   "content": "s"   }] } | |
| {   "id": "structural-00012",  "op": "WORDCHANGE",  "old": "anlyze",  "new": "analyse",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00051",  "op":"DEL",   "pos": 159,   "content": "anlyze"   },{   "id" : "edit-00052",  "op":"INS",   "pos": 159,   "content": "analyse"   }] } | |

### 3.1.3 Cambio di parola (WORDREPLACE)

Un'intera parola viene sostituita da un'altra. Qui la distinzione più importante è quella di parola (che forse è un termine sbagliato - forse sintagma?): rappresenta un'unità atomica di significato, ed è composta da un sostantivo con il suo eventuale articolo ed un eventuale aggettivo determinativo, oppure da un verbo e il suo eventuale verbo ausiliare. Ad esempio sono WORDREPLACE tutti i passaggi da una all'altra delle seguenti forme:

|  |  |  |
| --- | --- | --- |
| *Sostantivi* | *Verbi* | |
| word  words  a word  one word  the word  the words  some words  any word  any words  all words  every words | break  breaks  breaking  broke  broken  is breaking  has broken  will break  shall break  can break  may break | would break  should break  could break  might break  need to break  ought to break  must break  has to break  do break  does break  did break |

|  |  |
| --- | --- |
| **WORDREPLACE** | |
| *Sostituzione intera di una singola parola (la parola cambia).* | |
| {   "id": "structural-00013",  "op": "WORDREPLACE",  "old": "exclaimed",  "new": "said",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00054",  "op":"DEL",   "pos": 123,   "content": "exclaimed"   },{   "id" : "edit-00051",  "op":"DEL",   "pos": 123,   "content": "said"   }] } | |
| {   "id": "structural-00014",  "op": "WORDREPLACE",  "old": "a word",  "new": "some words",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00055",  "op":"DEL",   "pos": 152,   "content": "a"   },{   "id" : "edit-00056",  "op":"INS",   "pos": 152,   "content": "some"   },{   "id" : "edit-00057",  "op":"INS",   "pos": 161,   "content": "s"   }] } | |
| {   "id": "structural-00014",  "op": "WORDREPLACE",  "old": "a word",  "new": "some words",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00055",  "op":"DEL",   "pos": 152,   "content": "a word"   },{   "id" : "edit-00056",  "op":"INS",   "pos": 152,   "content": "some words"   }] } | |

### 3.1.4 Inserimenti e cancellazioni di testo (TEXTINSERT / TEXTDELETE)

|  |  |
| --- | --- |
| **TEXTINSERT** | **TEXTDELETE** |
| *Inserimento di una stringa di più parole diverse (escluso markup)* | *Eliminazione di una stringa di più parole (escluso markup)* |
| {   "id": "structural-00015",  "op": "TEXTINSERT",  "old": "",  "new": "neither more nor less",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00058",  "op":"INS",   "pos": 175,   "content": "neither more nor less"   }] } | {   "id": "structural-00016",  "op": "TEXTDELETE",  "old": "and ridiculous",  "new": "",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00059",  "op":"DEL",   "pos": 152,   "content": "and ridiculous "   }] } |

### 3.1.5 Sostituzioni di testo (TEXTREPLACE)

|  |
| --- |
| **TEXTREPLACE** |
| *Sostituzione di una stringa di più parole diverse (escluso markup)* |
| {   "id": "structural-10015",  "op": "TEXTREPLACE",  "old": "exactly one item",  "new": "one or more items",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-10058",  "op":"DEL",   "pos": 175,   "content": "exactly one item"   },{   "id" : "edit-10059",  "op":"INS",   "pos": 175,   "content": "one or more items"   }] } |

## 3.2 Operazioni sul markup

Sono operazioni sulla struttura quelle che avvengono tra più nodi di markup. A volte è modificato solo il markup, altre volte è modificato anche il testo intorno al markup. Ogni operazione è sempre accoppiata al suo inverso.

### 3.2.1 Modifiche senza effetto (NOOP)

Si tratta di modifiche meccaniche che, nella sintassi dei linguaggi di markup, non hanno alcun effetto concreto, come modifiche di ordine negli attributi di un elemento, o aggiunta o rimozione di whitespace in un nodo di testo.

|  |  |
| --- | --- |
| **NOOP** | |
| *Cambio non rilevante dentro al markup* | *Modifica non rilevante dentro al testo* |
| {   "id": "structural-10017",  "op": "NOOP",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-10060",  "op":"DEL",   "pos": 215,   "content": "<p id='x45' class='test'>"   },{   "id" : "edit-00060",  "op":"INS",   "pos": 215,   "content": "<p class='test' id='x45'>"   }] } | {   "id": "structural-10018",  "op": "NOOP",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-10061",  "op":"INS",   "pos": 255,   "content": "\n "   }] } |

### 

### 3.2.2 Inserimenti/cancellazione di nodi (INSERT/DELETE)

|  |  |
| --- | --- |
| **INSERT** | **DELETE** |
| *Inserimento di uno o più nodi nel documento, più, eventualmente, del testo prima e/o dopo ai nodi* | *Eliminazione di uno o più nodi nel documento, più, eventualmente, del testo prima e/o dopo ai nodi* |
| {   "id": "structural-00017",  "op": "INSERT",  "new": "Some text [...] more text",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00060",  "op":"INS",   "pos": 215,   "content": "<p>Some text</p><p>Some more text</p>"   }] } | {   "id": "structural-00018",  "op": "DELETE",  "old": "Some text [...] I don't like",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00061",  "op":"DEL",   "pos": 255,   "content": "<p>Some text to throw away</p><p>Some more text I don't like</p>"   }] } |
| {   "id": "structural-00019",  "op": "INSERT",  "new": "for instance [...] all are new.",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00062",  "op":"INS",   "pos": 275,   "content": "for instance: </p><p>Some text</p><p>Some more text</p><p>These all are new."   }] } | {   "id": "structural-00020",  "op": "DELETE",  "old": "for instance: [...] must go.",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00063",  "op":"DEL",   "pos": 307,   "content": "for instance: </p><p>Some text</p><p>Some more text</p><p>These all must go."   }] } |

### 3.2.3 Spostamenti (MOVE)

|  |  |
| --- | --- |
| **MOVE** | |
| *Spostamento di uno o più nodi da una parte all'altra del documento. Si noti che le due operazioni di rimozione e inserimento possono NON essere contigue nel tempo. L'ordine per edit meccanici le metterà nell'ordine temporale corretto, l'ordine per operazioni strutturali le metterà nell'ordine logico corretto. E' un move se e solo se il content del DEL e il content dell'INS coincidono e le due pos differiscono.* | |
| {   "id": "structural-00021",  "op": "MOVE",  "old": "Some text that goes away",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00064",  "op":"DEL",   "pos": 215,   "content": "<p>Some text that goes away</p>"   },{   "id" : "edit-00065",  "op":"INS",   "pos": 247,   "content": "<p>Some text that goes away</p>"   }] } | |

### 3.2.3 Abbassamenti/innalzamenti (WRAP/UNWRAP)

|  |  |
| --- | --- |
| **WRAP** | **UNWRAP** |
| *Annidamento di uno o più nodi con un nuovo nodo. Il contenuto non cambia, ma scende di un livello.* | *Rimozione di un nodo con promozione del suo contenuto al livello del nodo rimosso. Il contenuto non cambia, e sale di un livello.* |
| {   "id": "structural-00022",  "op": "WRAP",  "old": "the content of the node",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00065",  "op":"INS",   "pos": 535,   "content": "<i>"   },{   "id" : "edit-00066",  "op":"INS",   "pos": 540,   "content": "</i>"   }] } | {   "id": "structural-00023",  "op": "UNWRAP",  "old": "<i>the content of the node</i>",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00067",  "op":"DEL",   "pos": 535,   "content": "<i>"   },{   "id" : "edit-00068",  "op":"DEL",   "pos": 540,   "content": "</i>"   }] } |
| {   "id": "structural-00024",  "op": "WRAP",  "old": "the content of the node",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00069",  "op":"DEL",   "pos": 235,   "content": "the content of the node"   },{   "id" : "edit-00070",  "op":"INS",   "pos": 235,   "content": "<i>the content of the node</i>"   }] } | {   "id": "structural-00025",  "op": "UNWRAP",  "old": "<i>the content of the node</i>",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00070",  "op":"DEL",   "pos": 235,   "content": "<i>the content of the node</i>"   },{   "id" : "edit-00071",  "op":"INS",   "pos": 235,   "content": "the content of the node"   }] } |

### 3.2.4 Unioni/separazioni (JOIN/SPLIT)

|  |  |
| --- | --- |
| **JOIN** | **SPLIT** |
| *L'unione di due nodi fratelli e dello stesso tipo. Il nodo unito prende le caratteristiche del primo dei due nodi, e quelle del secondo nodo vengono perse.* | *La separazione di un nodo in due nodi fratelli dello stesso tipo. Entrambi i nodi prendono le caratteristiche del nodo unito.* |
| {   "id": "structural-00026",  "op": "JOIN",  "old": "First[...]paragraph.",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00072",  "op":"DEL",   "pos": 512,   "content": "</p><p>"   }] } | {   "id": "structural-00018",  "op": "SPLIT",  "old": "First[...]paragraph.",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00073",  "op":"INS",   "pos": 512,   "content": "</p><p>"   }] } |
| {   "id": "structural-00028",  "op": "JOIN",  "old": "First[...]paragraph.",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00072",  "op":"DEL",   "pos": 512,   "content": "<p>First paragraph.</p><p>Second paragraph.</p>"   },{   "id" : "edit-00073",  "op":"INS",   "pos": 512,   "content": "<p>First paragraph. Second paragraph.</p>"   }] } | {   "id": "structural-00029",  "op": "SPLIT",  "old": "First[...]paragraph.",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00074",  "op":"DEL",   "pos": 512,   "content": "<p>First paragraph. Second paragraph.</p>"   },{   "id" : "edit-00075",  "op":"INS",   "pos": 512,   "content": "<p>First paragraph.</p><p>Second paragraph.</p>"   }] } |

### 3.2.5 Sostituzioni (REPLACE)

|  |  |
| --- | --- |
| **REPLACE** | |
| *La sostituzione di un nodo con un altro senza modificare il contenuto. Dove appropriato vengono mantenuti gli attributi del vecchio nodo. Perchè sia considerata un replace, le modifiche meccaniche debbono garantire che la posizione è la stessa e il contenuto del nodo è lo stesso.* | |
| {   "id": "structural-00030",  "op": "REPLACE",  "old": "<li>The content of this list item</li>",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00076",  "op":"DEL",   "pos": 623,   "content": "<li>"   },{   "id" : "edit-00077",  "op":"DEL",   "pos": 645,   "content": "</li>"   },{   "id" : "edit-00078",  "op":"INS",   "pos": 623,   "content": "<p>"   },{   "id" : "edit-00079",  "op":"INS",   "pos": 645,   "content": "</p>"   }] } | |
| {   "id": "structural-00030",  "op": "REPLACE",  "old": "<li>The content of this list item</li>",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00076",  "op":"DEL",   "pos": 623,   "content": "<li>The content of this list item</li>"   },{   "id" : "edit-00078",  "op":"INS",   "pos": 623,   "content": "<p>The content of this list item</p>"   }] } | |

## 3.3 Approfondimenti sulle operazioni strutturali

### 3.3.1 Separazioni di operazioni meccaniche

I motori di diff non fanno distinzione tra modifiche al markup e al testo. Per questo motivo possono mettere insieme modifiche che a livello strutturale hanno spiegazioni diverse, solo perché sono contigue. Diventa allora importante essere in grado di spezzare una operazione meccanica in più operazioni, e inserirle in operazioni strutturali diverse. Tipicamente, queste operazioni avvengono agli estremi di nodi di testo, e non internamente. Ad esempio, si supponga di una modifica come segue:

|  |  |
| --- | --- |
| *Modifiche meccaniche da riorganizzare* | *Modifiche riorganizzate e separate tra più operazioni strutturali* |
| {  "id": "edit-10079",  "op": "DEL",  "pos": 80,  "content": "<p id='x45' class='test'>Testo iniziale."  },{  "id": "edit-10080",  "op": "INS",  "pos": 80,  "content": "<p class='test' id='x45'>Nuove parole."  } | {   "id": "structural-10031",  "op": "NOOP",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id": "edit-10079",  "op": "DEL",  "pos": 80,  "content": "<p id='x45' class='test'>"  },{  "id": "edit-10080",  "op": "INS",  "pos": 80,  "content": "<p class='test' id='x45'>"  }]  },{   "id": "structural-10032",  "op": "TEXTREPLACE",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id": "edit-10081",  "op": "DEL",  "pos": 94,  "content": "Testo iniziale."  },{  "id": "edit-10081",  "op": "INS",  "pos": 94,  "content": "Nuove parole."  }]  } |

### 3.3.2 Operazioni strutturali bilanciate e non bilanciate

Un concetto importante è quella di *operazione bilanciata*. Un'operazione è bilanciata se avviene su un sottoalbero del markup, a meno di nodi di testo parziali all'inizio e alla fine del markup. Noi ci occupiamo SOLO di operazioni bilanciate, cosicché non dobbiamo mai occuparci di ri-bilanciare l'albero dopo che l'operazione è stata effettuata. Un'operazione che avviene su una selezione sbilanciata deve essere ri-bilanciata prima di effettuare l'operazione separando in due o più operazioni bilanciate indipendenti la selezione complessiva.

|  |  |  |
| --- | --- | --- |
| *selezione bilanciata* | *selezione sbilanciata* | *Ribilanciamento della selezione* |
| <p>It was the <i>best of days</i>, it was the <i>worst of days</i>.</p> | <p>It was the <i>best of days</i>, it was the <i>worst of days</i>.</p> | <p>It was the <i>best of days</i>, it was the <i>worst of days</i>.</p> |
| <p>It was the <i>best of days</i>, it was the <i>worst of days</i>.</p> | <p>It was the <i>best of days</i>, it was the <i>worst of days</i>.</p> | <p>It was the <i>best of days</i>, it was the <i>worst of days</i>.</p> |

|  |
| --- |
| **DELETE su selezione sbilanciata** |
| {   "id": "structural-00031",  "op": "DELETE",  "old": "the best",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00079",  "op":"DEL",   "pos": 10,   "content": "the"   },{   "id" : "edit-00080",  "op":"DEL",   "pos": 17,   "content": "best"   }] } |
| {   "id": "structural-00032",  "op": "DELETE",  "old": "the best of days, it was the worst",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00081",  "op":"DEL",   "pos": 10,   "content": "the <i>best of days</i>, it was the"   },{   "id" : "edit-00082",  "op":"DEL",   "pos": 49,   "content": "worst"   }] } |

# 4 Operazioni semantiche

Sono operazioni semantiche tutte quelle (sequenze di) operazioni a cui è semplice dare un significato immediatamente comprensibile ad un essere umano. La distinzione è intuitiva, intrinsecamente imprecisa, per cui è soggetta ad interpretazione e può essere discutibile da persona a persona. Inoltre non è facile determinare un elenco definitivo e preciso di categorie utilizzabili.

Infine molte operazioni strutturali hanno un significato proprio semplice e ben descritto, per cui non ha senso raggrupparle in sovrastrutture. per questo motivo la caratterizzazione semantica viene usata solo a volte per raggruppare in sovrastrutture le operazioni strutturali, mentre negli altri casi viene considerata come una ulteriore modalità d descrizione di quella che è e rimane un'operazione strutturale. In questo caso parliamo più semplicemente di caratterizzazione o colorazione semantica di un'operazione strutturale.

Data la natura soggettiva ed interpretativa delle operazioni semantiche, lo scopo ultimo della caratterizzazione semantica delle operazioni strutturali è quella di fornire un ulteriore livello interpretativo, non quello di fornire una nuova realtà indiscutibile e definitiva, sopra le operazioni strutturali, che rimangono il livello massimo di caratterizzazione oggettiva delle operazioni che avvengono su un documento di testo. s

## 4.1 Colorazione semantica di operazioni strutturali

### 4.1.1 Correzione (FIX)

Sono correzioni quelle operazioni strutturali che risolvono un errore. Questo errore può essere grammaticale (un typo), strutturale (un paragrafo normale dentro ad una lista puntata) o concettuale (un'immagine senza didascalia) ma rappresenta sempre una transizione da uno stato considerabile come erroneo ad uno stato considerabile come corretto o più corretto. Qualunque operazione strutturale può essere considerata come una correzione. Ovviamente anche le edit wakes contengono una sequenza di correzioni, ma nello specifico contesto di riaggiustare progressivamente un errore (grammaticale) causato da una modifica vera precedente.

|  |  |  |
| --- | --- | --- |
| **FIX** | | |
| *Risoluzione di un errore* | | |
| {  "id": "semantic-00001",  "op": "FIX",  "new": "Some words are better than others",  "old": "Some words is better than others",  "items": {   "id": "structural-00033",  "op": "WORDREPLACE",  "by": "Fabio Vitali",  "timestamp": "2018-10-10T07:25:23.891Z",   "items": [{   "id" : "edit-00083",  "op":"DEL",   "pos": 215,   "content": "is"   },{   "id" : "edit-00084",  "op":"INS",   "pos": 215,   "content": "are"   }]  }  } | {  "id": "semantic-00002",  "op": "FIX",  "new": "<p>The content of this list item</p>",  "old": "<li>The content of this list item</li>",  "items": {   "id": "structural-00034",  "op": "REPLACE",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:25:23.891Z",   "items": [{   "id" : "edit-00085",  "op":"DEL",   "pos": 623,   "content": "<li>"   },{   "id" : "edit-00086",  "op":"DEL",   "pos": 645,   "content": "</li>"   },{   "id" : "edit-00087",  "op":"INS",   "pos": 623,   "content": "<p>"   },{   "id" : "edit-00088",  "op":"INS",   "pos": 645,   "content": "</p>"   }]  } } | |

### 4.1.2 Cambio stilistico (STYLE)

Sono cambi stilistici quelle operazioni strutturali che cambiano aspetto e/o contenuto del documento senza cambiarne il significato. Un cambio d'aspetto può essere una modifica di markup, oppure la sostituzione di una parola con un sinonimo, o anche un cambiamento di punteggiatura.

|  |  |  |
| --- | --- | --- |
| **STYLE** | | |
| *Modifica stilistica* | | |
| {  "id": "semantic-00003",  "op": "STYLE",  "new": "<b>some text</b>",  "old": "some text",  "items":{   "id": "structural-00024",  "op": "WRAP",  "by": "Fabio Vitali",  "timestamp": "2018-10-10T07:26:23.891Z",   "items": [{  "id" : "edit-00089",  "op":"INS",   "pos": 235,   "content": "<b>"   },{   "id" : "edit-00090",  "op":"INS",   "pos": 255,   "content": "</b>"   }]  } } | {  "id": "semantic-00004",  "op": "STYLE",  "new": "I bought a coffee",  "old": "I purchased a coffee",  "items": {   "id": "structural-00025",  "op": "REPLACE",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:27:27.891Z",   "items": [{   "id" : "edit-00091",  "op":"DEL",   "pos": 623,   "content": "purchased"   },{   "id" : "edit-00092",  "op":"INS",   "pos": 623,   "content": "bought"   }]  } } | {  "id": "semantic-00005",  "op": "STYLE",  "new": "John, Jack, and Bob",  "old": "John, Jack and Bob",  "items": {   "id": "structural-00026",  "op": "PUNCTUATION",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T07:28:28.891Z",   "items": [{   "id" : "edit-00093",  "op":"INS",   "pos": 623,   "content": ","   }]  } } |

### 4.1.3 Cambio di significato (MEANING)

In teoria, questa categoria è riservata alle sole operazioni che cambiano concretamente il significato del testo, come "Oggi è una bella giornata" e "Oggi ***non*** è una bella giornata". In pratica, questa è una categoria residuale: ogni modifica a cui non possiamo attribuire un'altra categoria viene interpretata come un cambio di significato, incluso "Oggi è una giornata di tempo sereno".

|  |
| --- |
| **MEANING** |
| *Modifica di significato* |
| {  "id": "semantic-00006",  "op": "MEANING",  "new": "Oggi non è una bella giornata",  "old": "Oggi è una bella giornata",  "items":{   "id": "structural-00027",  "op": "TEXTINSERT",  "by": "Fabio Vitali",  "timestamp": "2018-10-10T07:31:23.891Z",   "items": [{  "id" : "edit-00094",  "op":"INS",   "pos": 312,   "content": "non "   }]  } } |

## 4.2 Raggruppamenti semantici di operazioni strutturali

### 4.2.1 Edit wars (EDITWAR)

Wikipedia è molto precisa nel definire le edit war come segue: "An **edit war** occurs when editors who disagree about the content of a page repeatedly override each other's contributions." https://en.wikipedia.org/wiki/Wikipedia:Edit\_warring. Per noi un requisito ulteriore è che queste guerre avvengano ***sostanzialmente*** sulle stesse parti di testo e riportando il contenuto ***sostanzialmente*** allo stato precedente. Sono quindi modifiche ripetute, localizzate, e in cui il contenuto oscilla tra due o più stati diversi a cui ritorna con una certa regolarità. Non è necessario essere precisi nell'esigere la stessa locazione o lo stesso contenuto, ma circa nello stesso posto e circa lo stesso contenuto. Non è neanche un requisito quello che gli autori delle modifiche siano diversi: può essere sempre lo stesso autore che cambia ripetutamente idea sulla forma da dare ad un testo. Al minimo un'edit war è composta di due sostituzioni, tale per cui la prima modifica un testo e la seconda riporta il testo modificato alla condizione originaria. I campi new e old ci permettono di capire se, al momento, sta vincendo il testo originario (new == old) o quello modificato (new <> old). Altra cosa da notare, le singole operazioni strutturali di una edit war non sono né contigue né vicine nel tempo. Una edit war è in continuo aggiornamento. Un'eventuale ulteriore modifica, anche molto successiva alle precedenti, va accodata comunque alla edit war passata, e non costituisce una nuova edit war.

|  |
| --- |
| **EDITWAR** |
| *Guerra di modifiche* |
| {  "id": "semantic-00007",  "op": "EDITWAR",  "new": "Oggi è una bella giornata",  "old": "Oggi è una bella giornata",  "items":[{   "id": "structural-00014",  "op": "WORDREPLACE",  "old": "bella",  "new": "brutta",  "by": "Andrea Rossi",  "timestamp": "2018-03-10T09:25:23.891Z",   "items": [{   "id" : "edit-00095",  "op":"DEL",   "pos": 152,   "content": "bella"   },{   "id" : "edit-00096",  "op":"INS",   "pos": 152,   "content": "brutta"   }]  }, {  "id": "structural-00025",  "op": "WORDREPLACE",  "old": "brutta",  "new": "bella",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T11:45:11.891Z",   "items": [{   "id" : "edit-00106",  "op":"DEL",   "pos": 152,   "content": "brutta"   },{   "id" : "edit-00107",  "op":"INS",   "pos": 152,   "content": "bella"   }]  }  } |

### 4.2.2 Scia di modifiche (EDIT WAKE)

Una edit wake (scia di modifiche) è una modifica tipicamente "vera" (cioè semantica e contenutistica) a cui segue una scia di piccole modifiche causate dalla prima per ristrutturare e riorganizzare la frase in seguito ai problemi, anche banalmente grammaticali, causati dalla prima modifica. Ad esempio, la trasformazione da "John is going to his favorite restaurant with Mary" a "John and Mary are going to their favorite restaurant" è composta da una scia di modifiche che iniziano con l'inserimento della stringa "and Mary ", poi il cambio da "is" a "are", la trasformazione da "his" a "their", e infine la cancellazione di "with Mary".

|  |
| --- |
| **EDITWAKE** |
| *Modifica iniziale più una scia di modifiche correttorie come conseguenza* |
| {  "id": "semantic-00008",  "op": "EDITWAKE",  "new": "John and Mary are going to their favorite restaurant",  "old": "John is going to his favorite restaurant with Mary",  "items":[{   "id": "structural-00015",  "op": "TEXTINSERT",  "new": "and Mary ",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T11:13:23.891Z",   "items": [{   "id" : "edit-000108",  "op":"INS",   "pos": 215,   "content": "and Mary "   }]  },{   "id": "structural-00016",  "op": "WORDREPLACE",  "new": "are",  "old": "is",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T11:13:24.891Z",   "items": [{   "id" : "edit-00109",  "op":"DEL",   "pos": 224,   "content": "is"   },{   "id" : "edit-00110",  "op":"INS",   "pos": 224,   "content": "are"   }]  }, {  "id": "structural-000217",  "op": "WORDREPLACE",  "new": "their",  "old": "his",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T11:13:25.891Z",   "items": [{   "id" : "edit-00111",  "op":"DEL",   "pos": 221,   "content": "his"   },{   "id" : "edit-00112",  "op":"INS",   "pos": 221,   "content": "their"   }]  },{  "id": "structural-00218",  "op": "TEXTDELETE",  "old": "with Mary",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T11:13:28.891Z",   "items": [{   "id" : "edit-00112",  "op":"DEL",   "pos": 230,   "content": "with Mary"   }]  }]  } |

### 4.2.3 Catene di modifiche (EDIT CHAIN)

Una edit chain (catena di modifiche) è una serie di modifiche uguali o molto simili attivate spesso in maniera automatica, come un cerca e sostituisci globale o un cambiamento strutturale importante (una serie di paragrafi trasformati in lista).

|  |
| --- |
| **EDITCHAIN** |
| *Modifiche ripetute e molto simili.* |
| {  "id": "semantic-00009",  "op": "EDITCHAIN",  "new": "John (3 times)",  "old": "Jack",  "items":[{   "id": "structural-00019",  "op": "WORDREPLACE",  "new": "John",  "old": "Jack",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T11:15:24.891Z",   "items": [{   "id" : "edit-00120",  "op":"DEL",   "pos": 224,   "content": "Jack"   },{   "id" : "edit-00121",  "op":"INS",   "pos": 224,   "content": "John"   }]  },{   "id": "structural-00020",  "op": "WORDREPLACE",  "new": "John",  "old": "Jack",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T11:15:24.891Z",   "items": [{   "id" : "edit-00122",  "op":"DEL",   "pos": 254,   "content": "Jack"   },{   "id" : "edit-00123",  "op":"INS",   "pos": 254,   "content": "John"   }]  },{   "id": "structural-00021",  "op": "WORDREPLACE",  "new": "John",  "old": "Jack",  "by": "Fabio Vitali",  "timestamp": "2018-03-10T11:15:24.891Z",   "items": [{   "id" : "edit-00124",  "op":"DEL",   "pos": 316,   "content": "Jack"   },{   "id" : "edit-00125",  "op":"INS",   "pos": 316,   "content": "John"   }]  }]  } |

# 5 Modifiche alle versioni precedenti

## 5.1 Modifiche tra la versione senza numero (1.0) e la 1.1

* Introdotto un sistema di numerazione delle versioni.
* Aggiunta l'operazione strutturale NOOP
* Aggiunta l'operazione strutturale TEXTREPLACE
* Aggiunta la sezione 3.3.1 (*Separazione di operazioni meccaniche*)
* Aggiunta la sezione 5 (*Modifiche alle versioni precedenti*)

## 