

Hibernate Search

Google Your Domain Model



Pär Wenåker | par.wenaker@callistaenterprise.se | Cadec 2010-01-20



Agenda

- What's the problem to solve?
- Apache Lucene
- Hibernate Search





Search functionality



Search Form

Find issues that have...

all these words:

this exact phrase:

one or more of these words:

 OR OR

But don't show issues that have...

these unwanted words:

▼ [More](#)

Project:

Sprint:

Tags:

(Enter here the list of comma separated tags e.g.: Bug, Ops, Feature)

Owner:

Status:

Priority:

Estimate:

Remaining:

See [help](#) for more [search options](#)





Simpler, like Google You know...

Google™
Sverige

[Avancerad sökning](#)
[Språkverktyg](#)

Google-sökning Jag har tur


Sök: webben sidor på svenska sidor från Sverige



What is requested?

- Show most relevant results first.
- Handle typos.
- Find words with similar meaning.
- Find words from same family.





SQL
...where bla bla
LIKE '%Stuff%'



Leads to complex SQL

with x(s, ind) as

```
( select sud, instr( sud, ' ' )
```

```
from ( select '53 7 6 195 98 6 8 6 34 8 3 17 2 6 6 28 419 5 8 79' sud from dual )
```

```
union all
```

```
select substr( s, 1, ind - 1 ) || z || substr( s, ind + 1 ) , instr( s, ' ', ind + 1 )
```

```
from x , ( select to_char( rownum ) z from dual connect by rownum <= 9 ) z
```

```
where ind > 0
```

```
and not exists ( select null
```

```
from ( select rownum lp from dual connect by rownum <= 9 )
```

```
where z = substr( s, trunc( ( ind - 1 ) / 9 ) * 9 + lp, 1 )
```

```
or z = substr( s, mod( ind - 1, 9 ) - 8 + lp * 9, 1 )
```

```
or z = substr( s, mod( trunc( ( ind - 1 ) / 3 ), 3 ) * 3
```

```
+ trunc( ( ind - 1 ) / 27 ) * 27 + lp
```

```
+ trunc( ( lp - 1 ) / 3 ) * 6
```

```
, 1 )
```

```
)
```

```
)
```

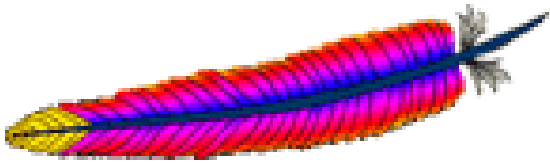
```
select s from x where ind = 0
```





Apache
Lucene





- A Java fulltext search engine library

Indexing

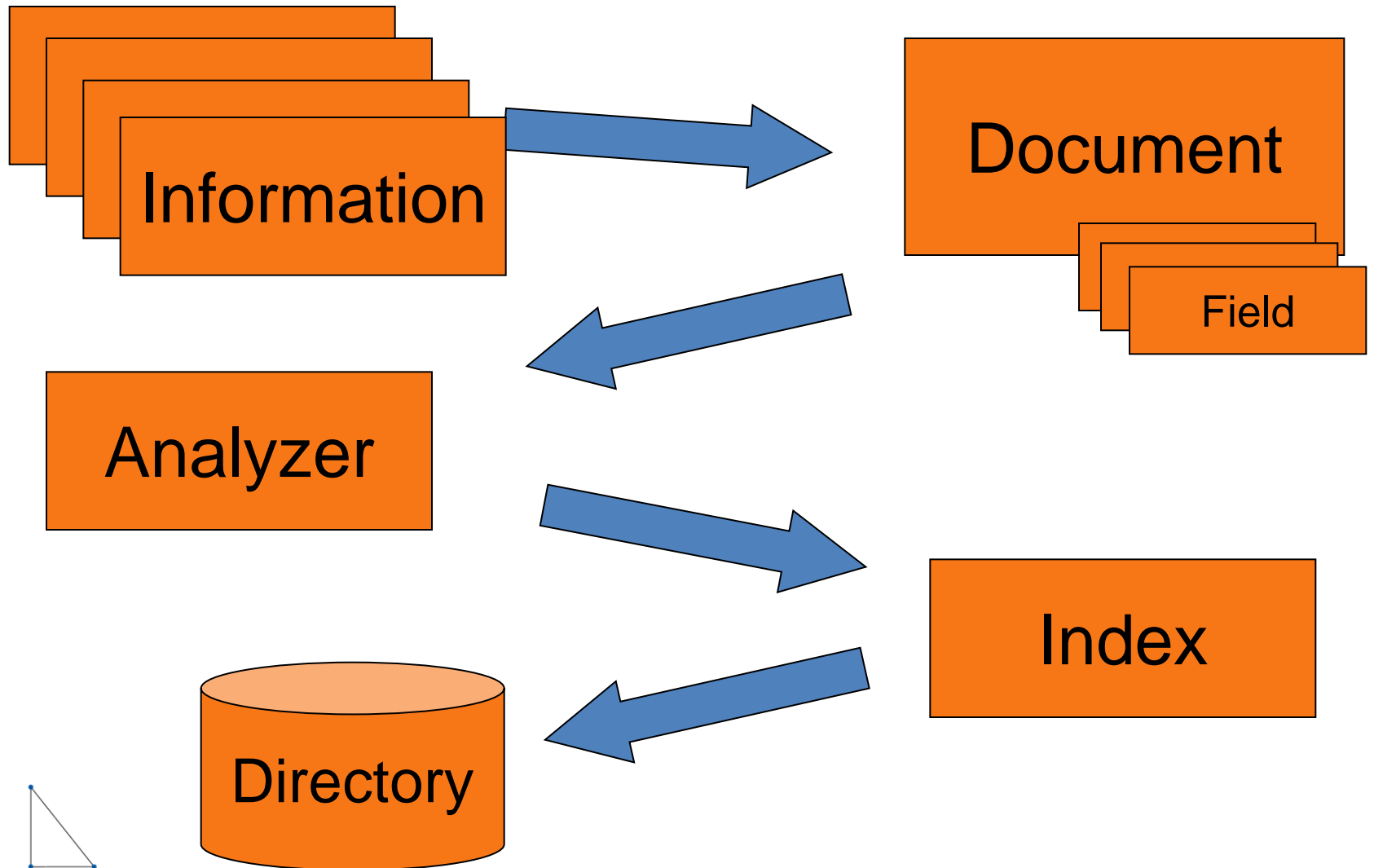
Searching



Indexing



Indexing



Standard Analyzer Example

Look that's a droid

StandardTokenizer

Look

that's

a

droid

StandardFilter

Look

that

a

droid

LowerCaseFilter

look

that

a

droid

StopFilter

look

droid



Searching



Searching

- Formulate a Query
- Find documents matching the query.
- Rank the matches by relevance.



Lucene Query Syntax

- artist:coldplay
- artist:coldpay~
- title:"speed sound" ~1
- +artist:coldplay +title:vida
- artist:coldplay title:la^2.0
- year:[1997 TO 2007]



Searching

```
Searcher searcher = new IndexSearcher(FSDirectory.getDirectory("index"));
```

```
QueryParser parser = new QueryParser("artist", analyzer);
```

```
Query query = parser.parse("killers coldplay");
```

```
TopDocs docs = searcher.search(query, 10);
```

```
int hits = docs.totalHits
```

```
ScoreDoc[] hits = docs.scoreDocs;
```



Programmatic API

- TermQuery
- PhraseQuery
- WildcardQuery
- RangeQuery
- FuzzyQuery
- BooleanQuery



Searching

```
Term t1 = new Term("artist", "coldplay");  
Term t2 = new Term("artist", "killers");
```

```
Query q1 = new TermQuery(t1);  
Query q2 = new TermQuery(t2);  
q1.setBoost(2.0);
```

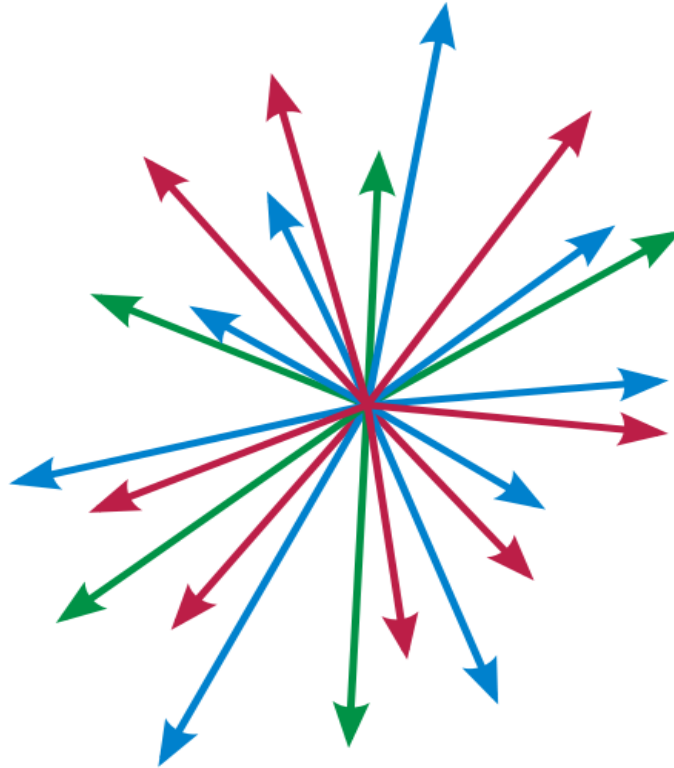
```
Query query = new BooleanQuery();  
query.add(q1, BooleanClause.Occur.MUST);  
query.add(q2, BooleanClause.Occur.SHOULD);
```

```
TopDocs docs = searcher.search(query, 10);
```



Document scoring

$$\sum_{t \text{ in } q} (tf(t \text{ in } d) \times idf(t)^2 \times boost(t.\text{field in } d) \times lengthNorm(t.\text{field in } d)) \times coord(q, d) \times queryNorm(q)$$



Aproximative Search

- FuzzySearch
- N-gram
 - Cal,all,lli,lis,ist,sta
- Phonetic search
- Synonyms



Concurrency

- Only ONE IndexWriter allowed at one time.
- Multiple IndexReaders allowed concurrently.
- Lucene uses file locks by default.



Demo Luke

The screenshot shows the Luke Lucene Index Toolbox v 0.9.9 (2009-09-30) interface. The search expression is "artist:hsajashdjaskhdhj". The analyzer is set to "org.apache.lucene.analysis.KeywordAnalyzer" and the default field is "_hibernate_class". The search results section shows "No Results" and a table with columns: #, Score, Doc. Id, _hibernate_cl:artist, id, title, tracks.name, year. The index name is "/Users/par....callista.cadec.model.Album".

Enter search expression here:

Analysis | QueryParser | Similarity | Collector

Analyzer to use for query parsing:
NOTE: use fully-qualified class name here. Default field:
org.apache.lucene.analysis.KeywordAnalyzer | _hibernate_class
Optional constructor argument:

Query details: Update Explain structure
artist:hsajashdjaskhdhj
Parsed
Rewritten


Last search time: 449 us Search repeat 1 times.

Results: (Hint: Double-click on results to display all fields) Explain Delete 0 doc(s) 0-19

#	Score	Doc. Id	_hibernate_cl:artist	id	title	tracks.name	year
No Results							

Index name: /Users/par....callista.cadec.model.Album





But I have a
RDBMS and I
am using JPA...



Mismatches

- Structural mismatch
- Synchronization mismatch
- Retrieval mismatch





Hibernate
Search



Mapping

@Entity

@Indexed

```
public class Album {
```

```
    @Id
```

```
    @GeneratedValue
```

```
    @DocumentId
```

```
    private Integer id;
```

```
    @Field(index = Index.TOKENIZED)
```

```
    private String artist;
```

```
    @Field(index = Index.TOKENIZED, store=Store.YES)
```

```
    private String title;
```

```
    @Field(index = Index.UN_TOKENIZED, store=Store.YES)
```

```
    private String year;
```

```
    @OneToMany(cascade=CascadeType.ALL)
```

```
    @JoinColumn(name="album_id")
```

```
    @IndexedEmbedded
```

```
    private Set<Track> tracks = new HashSet<Track>();
```



Define a custom analyzer

```
@AnalyzerDef(  
    name="standard",  
    tokenizer =  
        @TokenizerDef(factory = StandardTokenizerFactory.class),  
    filters = {  
        @TokenFilterDef(factory = StandardFilterFactory.class),  
        @TokenFilterDef(factory = LowerCaseFilterFactory.class),  
        @TokenFilterDef(factory = StopFilterFactory.class,  
            params = @Parameter(name="words", value = "stopwords.txt")),  
    }  
)  
@Indexed  
public class Album {
```



Configuration

```
<persistence xmlns="http://java.sun.com/xml/ns/persistence"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://java.sun.com/xml/ns/persistence
http://java.sun.com/xml/ns/persistence/persistence_1_0.xsd version="1.0">
  <persistence-unit name="pu" transaction-type="RESOURCE_LOCAL">
    <properties>
      <property name="hibernate.connection.driver_class" .... />
      <property name="hibernate.connection.password" value="" />
      <property name="hibernate.connection.url" value="..." />
      <property name="hibernate.connection.username" value="..." />
      <property name="hibernate.dialect" value="org.hibernate.dialect.MySQLDialect" />
      <property name="hibernate.search.default.indexBase" value="./index" />
    </properties>
  </persistence-unit>
```



Synchronization

- Transparently handles indexing
- Lucene Documents are read only (delete/create)
- Apply indexing in an optimal way
- Index is NOT a XA resource



Retrieval

- `javax.persistence.EntityManager`
- `javax.persistence.Query`

- `org.hibernate.search.jpa.FullTextEntityManager`
- `org.hibernate.search.jpa.FullTextQuery`



Programming searching

```
EntityManager em = ...
```

```
// Get a FullTextEntityManager
```

```
FullTextEntityManager ftEm =  
    org.hibernate.search.jpa.Search.getFullTextEntityManager(em);
```

```
// Create a Query
```

```
Query query = ....
```

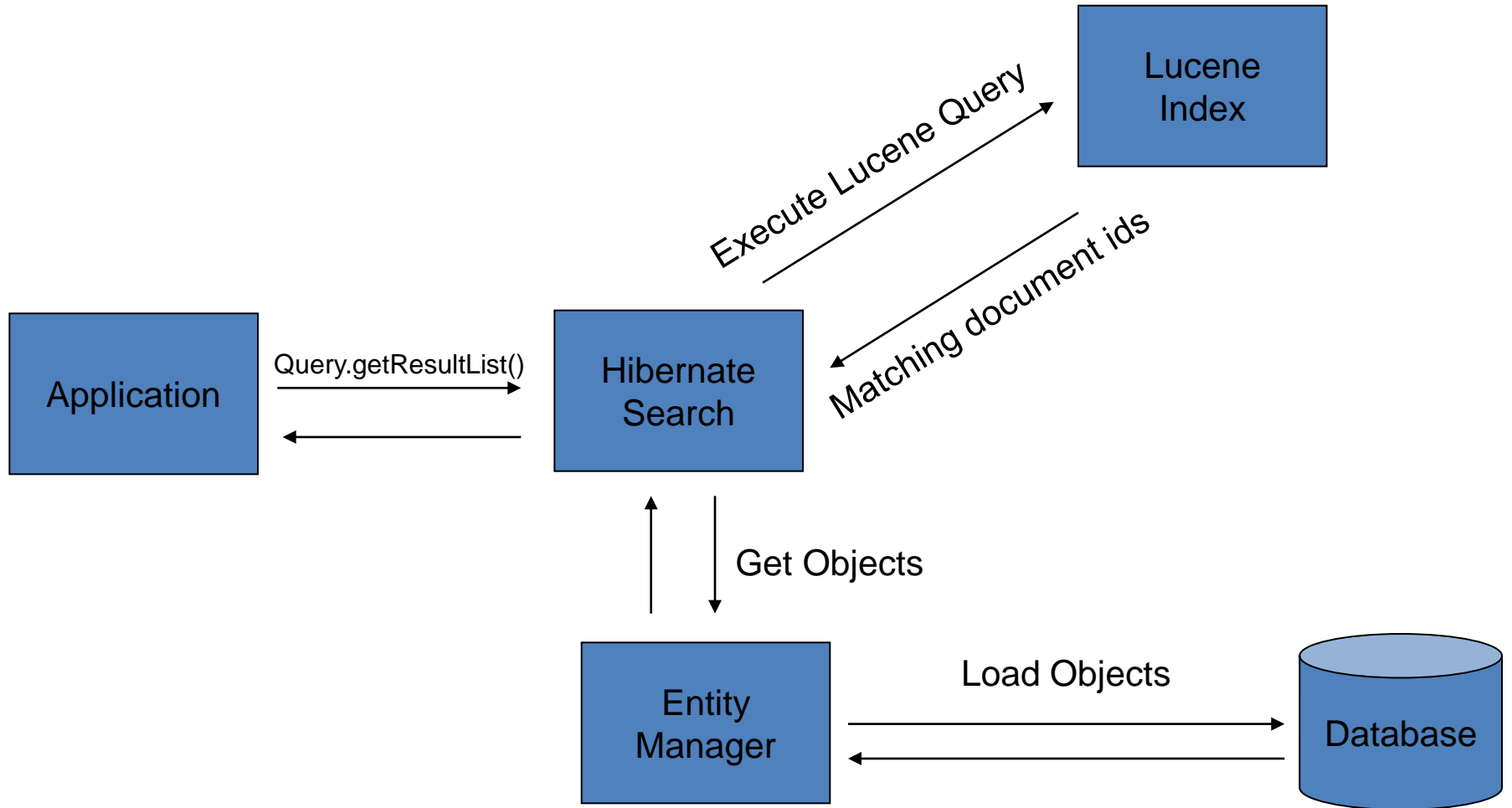
```
FullTextQuery ftQuery = ftEm.createFullTextQuery(query, Album.class);
```

```
// Execute the
```

```
QueryList<Album> results = ftQuery.getResultList();
```



Search



Example from real life...



[Startsidan](#)

[Sjukdomar och besvär](#)

[Undersökningar och behandlingar](#)

[Hitta rätt i vården](#)

[Regler och rättigheter](#)

[Mina tjänster](#)

/ Webbisar

Webbisar

Här kan föräldrar lägga in bilder på sina nyfödda bebisar. Detta är en gratis och reklamfri webbtjänst som ägs och drivs av Västra Götalandsregionen. För att skapa ett konto behöver du ha en [e-legitimation](#).

[Lägg till / Redigera webbis](#)

Exempel: 20090211 +Erik -Johan Svensson Östra

Oscar född i fredags, Linda och Per



Tid: 05:26
Vikt: 4220g
Längd: 54cm

Föräldrar:
Linda Wall
Per Wall

Från: Borås
Sjukhus: SAS Borås

Gustav född 6 januari, Niklas och Maria



Tid: 17:37
Vikt: 3770g
Längd: 54cm

Föräldrar:
Niklas Andersson
Maria Bergqvist

Från: Borås
Sjukhus: SAS Borås

Figuren född 3 januari, Mariana och Mikael



Tid: 17:18
Vikt: 2980g
Längd: 51cm

Föräldrar:
Mariana Nilsson
Mikael Nilsson

Från: Björlanda
Sjukhus: Mölndals sjukhus

Eddie född 2 januari, Jessica och Pontus



Tid: 20:10
Vikt: 3370g
Längd: 49cm

Föräldrar:
Jessica Stenström
Pontus Löfgren

Från: Trollhättan
Sjukhus: NÄL

Agnes Eklöv född 2 januari, Sofia och Erik



Tid: 15:14
Vikt: 3650g
Längd: 52cm

Föräldrar:
Sofia Eklöv
Erik Eklöv

Från: Pålvelund, Göteborg
Sjukhus: Mölndals sjukhus

Mio född 2 januari, Sofia och Morgan



Tid: 08:27
Vikt: 4480g
Längd: 52cm

Föräldrar:
Sofia Holmgren
Morgan Scammel

Från: Limmared
Sjukhus: SAS Borås

[Fler webbisar](#)

Sök vårdmottagning

Mottagningens namn

Sök direkt på namn

Vilken typ av vård?

Alla typer av vård

Var vill du hitta vård?

Välj kommun

Lista Karta

Senast uppdaterad: 2009-01-28 Publicerad av: Vårdredaktionen vardredaktion@vgregion.se

[Om Webbplatsen](#) [Om Cookies](#)



Informationsansvarig: Margreth Jonson Lekare. Denna sida är en del av www.vgregion.se

/ Webbisar

Webbisar

Här kan föräldrar lägga in bilder på sina nyfödda bebisar. Detta är en gratis och reklamfri webbtjänst som ägs och drivs av Västra Götalandsregionen. För att skapa ett konto behöver du ha en [e-legitimation](#).

[Till webbis-startsidan](#)

[Lägg till / Redigera webbis](#)

Exempel: 20091011 +Erik-Johan Svensson Östra

3 träffar

[Helene Gran och Jens Gran, Brastad, NÄL](#)

Emma, 2009-10-14 23:31, 4430g, 51cm

[Emma Fahlander och Michael Sundén, Borås, SÄS Borås](#)

Isac 2010 01 01, 2010-01-01 13:26, 3120g, 47cm

[Emma Torsson och Johan Andersson, Hisings backa, Östra sjukhuset](#)

Livia, 2009-11-13 07:05, 3220g, 52cm

Sök vårdmottagning

Mottagningens namn

Vilken typ av vård?

Var vill du hitta vård?

Lista Karta



/ Webbisar

Webbisar

Här kan föräldrar lägga in bilder på sina nyfödda bebisar. Detta är en gratis och reklamfri webbtjänst som ägs och drivs av Västra Götalandsregionen. För att skapa ett konto behöver du ha en [e-legitimation](#).

[Till webbis-startsidan](#)

[Lägg till / Redigera webbis](#)

Exempel: 20091011 +Erik -Johan Svensson Östra

28 träffar

[Maria Henriksson och Peter Flodin, Göteborg, Östra sjukhuset](#)

Estelle, 2009-12-25 02:32, 3500g, 49cm

[Maria Novoa Tafuri och Daniel Novoa, Göteborg, Östra sjukhuset](#)

Ana-Lucia, 2009-12-06 02:14, 3570g, 49cm

[Niklas Andersson och Maria Bergqvist, Borås, SÄS Borås](#)

Gustav, 2010-01-06 17:37, 3770g, 54cm

[Ingemar Bredström och Maria Bredström, Orust, NÄL](#)

Flicka, 2009-12-26 17:48, 4600g, 52cm

[Maria Rydin och Bruno Andersson, Sandared, SÄS Borås](#)

Signe, 2009-11-07 09:17, 3930g, 52cm

[Maria Karlsson och Glenn Karlsson, Vänersborg, NÄL](#)

Arwin, 2009-09-06 23:49, 4325g, 54cm

[Amanda Gruneau och Daniel Lobelius Gruneau, Göteborg, Östra sjukhuset](#)

Lova Gruneau, 2010-01-02 03:04, 2535g, 45cm

[Ida Ericsson och Martin Goding, Göteborg, Östra sjukhuset](#)

Juni, 2009-12-31 05:32, 3690g, 50cm

[Fler sökresultat](#)

Sök vårdmottagning

Mottagningens namn

Vilken typ av vård?

Var vill du hitta vård?

Lista Karta

Hibernate Search

- Enables transparent fulltext indexing
- Enables fulltext querying á la JPA/Hibernate
- Lets you focus on the query

Source Webbisar

<http://code.google.com/p/oppna-program-webbisar/>

Vård i Västra Götaland

VÄSTRA GÖTALANDSREGIONEN 1177

Startsidan | Sjukdomar och besvär | Undersökningar och behandlingar | Hitta rätt i vården | Regler och rättigheter | Mina tjänster







/ Webbisar

Webbisar

Här kan föräldrar lägga in bilder på sina nyfödda bebisar. Detta är en gratis och reklamfri webbtjänst som ägs och drivs av Västra Götalandsregionen. För att skapa ett konto behöver du ha en [e-legitimation](#).

[Lägg till / Redigera webbis](#)


Exempel: 20090211 +Erik -Johan Svensson Östra

Oscar född 1 fredags, Linda och Per  <p>Tid: 05:26 Vikt: 4220g Längd: 54cm</p> <p>Föräldrar: Linda Wall Per Wall</p> <p>Från: Borås Sjukhus: SAS Borås</p>	Gustav född 6 januari, Niklas och Maria  <p>Tid: 17:37 Vikt: 3770g Längd: 54cm</p> <p>Föräldrar: Niklas Andersson Maria Bergqvist</p> <p>Från: Borås Sjukhus: SAS Borås</p>
Figuren född 3 januari, Mariana och Mikael  <p>Tid: 17:18 Vikt: 2980g Längd: 51cm</p> <p>Föräldrar: Mariana Nilsson Mikael Nilsson</p> <p>Från: Björlanda Sjukhus: Mölndals sjukhus</p>	Eddie född 2 januari, Jessica och Pontus  <p>Tid: 20:10 Vikt: 3370g Längd: 49cm</p> <p>Föräldrar: Jessica Stenström Pontus Löfgren</p> <p>Från: Trollhättan Sjukhus: NÄL</p>
Agnes Eklöv född 2 januari, Sofia och Erik  <p>Tid: 15:14 Vikt: 3650g Längd: 52cm</p> <p>Föräldrar: Sofia Eklöv Erik Eklöv</p> <p>Från: Pålvelund, Göteborg Sjukhus: Mölndals sjukhus</p>	Mio född 2 januari, Sofia och Morgan  <p>Tid: 08:27 Vikt: 4480g Längd: 52cm</p> <p>Föräldrar: Sofia Holmgren Morgan Scammel</p> <p>Från: Limared Sjukhus: SAS Borås</p>

[Fler webbisar](#)

Sök vårdmottagning

Mottagningens namn

Vilken typ av vård? 

Var vill du hitta vård?

Lista Karta

Senast uppdaterad: 2009-01-28 Publicerad av: Vårdredaktionen vardredaktion@vgregion.se Om Webbplatsen Om Cookies

VÄSTRA GÖTALANDSREGIONEN Informationsansvarig: Margreth Jonson Lekare. Denna sida är en del av www.vgregion.se

Attributions



by gerlos
<http://www.flickr.com/photos/gerlos/3119891607>



by llawliet
<http://www.flickr.com/photos/llawliet/2547595587>



by WILLGT09
<http://www.flickr.com/photos/williambrawley/4192206382>



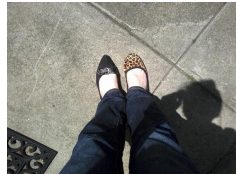
by Hans Dekker
<http://www.flickr.com/photos/hansdekker/398782008>



by Bart van de Blezen
<http://www.flickr.com/photos/bartelomeus/4184705426>



by Stéfan
<http://www.flickr.com/photos/st3f4n/3951143570>



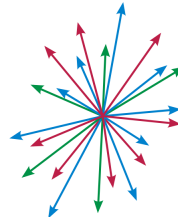
by Irna slutsky
<http://www.flickr.com/photos/irinaslutsky/1043756538>



by quinn.anya
<http://www.flickr.com/photos/quinnanya/3250989584>



by woodleywonderworks
<http://www.flickr.com/photos/wwworks/2472230611>



by Ethan Hein
<http://www.flickr.com/photos/ethanhein/2260878305>

