Unassigned jobs

< GENERAL>

- Upgrade-script
- make TSEP indepedent from MySQL -> use ADODB instead. Todo: Make IP sort algorithm, check how to acomplish PHP boolean search
- correct some div class names to be matching more what they do
- delete obsolete classes from CSS file
- update configuration.php-/indexer.php-Item-Texts and Item-Help (delete obsolete information example: “zb must not have / in the end”, “only true/false”

<output>

PRIO 25

TSEP should be able to mark each search term in the search results with a different DIV/SPAN tag. Right now TSEP marks all search terms with the same tag. TSEP should mark every each search term with a different tag class.
Example: If the user searches for
apple tree
TSEP should mark these words in the search results individually:
apple would be inside a tag “.tsepProject .highlight .term_1”,
tree would be inside a tag “.tsepProject .highlight .term_2”.
TSEP should number (create) those tags automatically, so that it does not matter how many search terms have been used. The admin can then create the css styles like he wishes – either all the same, groups, or for example totally different styles.

PRIO 5

complete output with templates and variables for tsep content: This means, we want to use a templating engine to output all TSEP files which the administrator might want to fit into hit own layout. This concerns the search input and output (results) as well as the search-tipps. The content of the pages needs to be broken down into pieces that we can throw into a templating engine. We found that SMARTY if to much for our purpose, but that the SMARTPHP engine should work just fine. http://www.smartphp.net/

</output>

Security

PRIO 7

Suggestion from Daniel S.: I was trying to hack the tsep myself, and managed to do it making the following changes. I know very little about sql injection, but gather with magic quotes on (which is default on recent php installs, and certainly mine) the best defence is simply to escape strings with ' ".$string." ' (spaces inserted to make it easier to read). So I got my version working by:

in search.php

remove line "deslash($q);" line 58

remove line "reslash($searchFor);" line 301

ammend line "$lSrchInList = preg_replace("/[-+\(\)<>\"\']/", "", $lSrchInList);" line 305 to

"$lSrchInList = preg_replace("/[-+\(\)<>\"]/", "", $lSrchInList);"

ammend line 308 to .... WHERE stopword IN ('".$lSrchInList."') ORDER BY stopword";

in cleanstring.php

remove line "$wild = str_replace("\"", "", $wild);" line 45

/security

</ GENERAL>

<indexer>

<improve indexer output when indexing>
PRIO 9

> similar to indexoverview show only title, URL, and the number of indexed words. No profiles should be shown there as there could be many. This results in a very small output of only one line per indexed file.
> As in indexoverview :
> 1. a click on the title (if no title, show URL at this place) takes the user to Indexedit of that page
> 2. a click on the URL takes the user to the page itself

</improve indexer output when indexing>

- in indexer profile: read robots.txt and follow the rules set there or ignore them?
- automatically read domain to index in the indexer but let the admin change this setting (should be finished by now)
- **PRIO 8** index pages which have this meta: <META NAME="robots" CONTENT=" noindex"> or just leave those pages out
- Add an extra field to the database: Additional information of a page. Here the admin can enter extra information (similar top the internal info field) BUT in a basket can be be defined if the addtional information (of course it can be left empty) will be a) shown alone, b) shown together with the indexed page contend c) not at all.
- **PRIO 30** try to see what a page returns – if it is an error (404 for example) or “non-ascii”-garbage, we should not add the page to the database but show the user an error and explain what happens

</indexer>

<indexedit>

- **PRIO 15** indexedit: make it possible to change attached profiles of the pages

</indexedit>

<logs>

PRIO 10

I think the idea of giving the user some options in the logview is quite good.

/admin/logview.php

Ideas for handling:
The user can filter on the following criterias - actually the logview should _**always be filtered**_ on this criteria but the user can change it:

Of course the admin should be able to filter the time:
- from starting date (Format example: YYYY/MM/DD) (including) [2000/01/01 default]
- until ending date (Format example: YYYY/MM/DD) (including) [now default]

for all criterias filters should exist in the logview. Examples:
- that are type 1 (searchterm) or 2 (resultclick) or both [both default]
- and contain the following characters [* for this does not matter] which are produced by this ipaddress
- range start: 000.000.000.000 [default: 000.000.000.000, we write 000.000.000.000 to db when ip is not logged]
- range end: 255.255.255.255 [default: 255.255.255.255]

A button "update" to create a list that matches the criterias

A deletion of records it _**not**_ needed, useful nor wished: Deleting records will change any statistics we might create.

IP resolving:
- resolve ip addresses. Additional column + sorting possible here a well. Also another field in the filter criteria: "domain contains:" .... (user could enter .org or microsoft.com ... [default: *]
- There could also be a checkbox above the column "resolved address" which either enables resolving of disables it. If disabled, show "not resolved", if it could not be resolved "unable to resolve"

Idea for statistics:
- Count how often the same thing happened: How often has been searched for a term, how often has been clicked on a link. The output would be similar to the logview now but show the count number. This might need to be done in an extra file (page).
- Add an entry (new column in database table) in the log which contains the number of pages that have been returned by the search terms used (at the time the search was done)

Ideas for searched stopwords:
- show in log when the searched term was a stopword (at that time / still is a stopword = 2 different symbols maybe), maybe mark it with character which can not be searched anyways.
  Maybe ^ or ° - the stopword can be between the character. Example: If the stopword is “links” it would show up in logview as ^links^ or °links° (bad idea, read next paraprag, that is much better!)
  Better would be to introduce a new column in the database to show that the search word was a stopword at that time (boolean field type). This coulum should be shown next to the search entry colum. Also there should be another column that shows the state of the words at the time the logview is opened: What searchterms from the DB are right now in the stopword list

</logs>

<indexoverview>

- **PRIO 9** add filter possibility to indexoverview (i.e.: drop down field with all profiles and "- all -", the only one chosen will be shown)
- fold-unfold mechanism in indexoverview (and generally in url-list-areas)

</indexoverview>

<stopwords>

PRIO 10

<create stopword from index>Updates 2005-01-27 ON, added 2005-01-26, by Olaf Noehring:

- Stopwords overview page: All stopwords are shown on this page sorted by their occurence in the pages (life lookup!). The number of occurences will be displayed next to each stopword. If user clicks on the number – show the pages in which the stopword appears (another life search might be needed for this) This list could (should) be the indexoverview – but only containing the "filtered" (see PRIO 9 add filter possibility to indexoverview) pages
- Also next to each stopword will be a checkbox. The admin can check the checkbox and press a "submit" button to remove all checked stopwords from the stopwords list. This enables us to create individual stopword lists for each website very fast.
- The admin will have also the possiblity to enter a number (e.g. "100) to mark the first 100 entries automatically.
- Also the admin can click on a link "all stopwords to this stopword" (next to each stopword) to check all of the stopwords above that link. (Java script)
- This "Create stopwords from index" page should be an extra page in the admin area
- The admin can enter a number in a form on the page example "50". Afterwards the 50 (in this example) most occuring words of all indexed pages will be added as stopwords.
- add sorting by appearance / and word (asc/desc)

</create stopword from index>

</stopwords>

< search>

- **Prio 30** Introduce weighting of HTML elements. See "HTML Weighting details" section for details
- **Prio 40** Search in results: The user may want to search something in a resulted search.
- **Prio 47** add possibility for enduser to search in some profiles only (filter criteria) needs administrator settings: Are some profiles always to be searched?, Is choosing the profiles allowed to the user?+ stopwords: include a text list of make a textfield to copy and paste a list into. EXTEND this to "profile baskets"
- **Prio 15** change the icon that can represent the rank: Show not 1 symbol for 1 hit of a search term, but calculate how many terms are at most in one file – this is 100% - all others should be rated by percentage. The Admin has to set how many symbols equal that 100%. Example: 100% = 10 symbols. The hit with most terms (rank 1) shows 10 symbols, a hit with only half of the terms in the file shows 5 symbols.

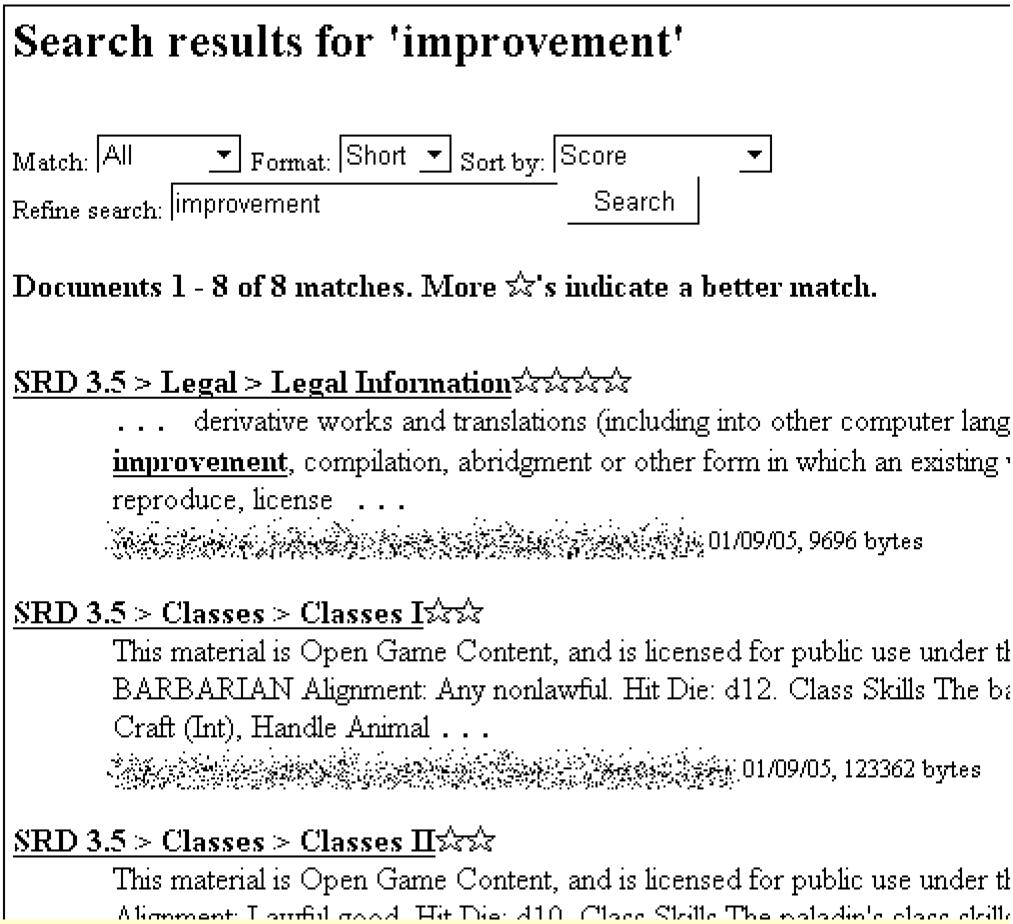
- **Prio 85** introduce new search command “near” - to add the restriction that to words / terms must not be further apart from each other than “z” words (z to be defined by the admin)
- **Prio 35** make it possible to show “the last X searchterms were: ...” then list what the people have been searching, and maybe the time when that was
  The number of terms shown (X) is 0 for: “Do not show that at all” to something >0.The administrator can define the number in configuration.php. Also he can define if the time/date should be shown
  The number of resulting pages that were returned by that search could be shown next to the search term in brackets: “links [9]” for example. If this shall be shown at all and the form of the brackets could be defined by the admin too. (“Show '[' before number of hits, show ']' after number of hits” in the case of my example.) To show the number we would need a new field in the log table which stores the number of resulting pages OR: do a life search for all search terms – but this might take way too long, especially if the X (above) is a high number. Nevertheless the last solution would produce better results.

<htaccess issue>

> Es geht darum, einem Usern, die Berechtigungen zu
> (htaccess)-gesperrten Bereichen haben, mehr als Suchergebnis zeigen, als anderen.
> Nur wie willst Du das beim Suchstart unterscheiden?
>
> eben, beim Suchstart muss über den Aufruf von TSEP schon ein Parameter an TSEP
> übergeben werden der sagt welcher Basket bzw. welche Baskets zu nutzen ist/sind.
> Man könnte den Admin z.B. eine ID übergeben lassen, die des Baskets. Dann müssen
> aber Baskets in Baskets vorkommen können. (Baum)
> So sind ganz unterschiedliche Suchfelder auf einer Domain möglich - was nach
> deiner Vorstellung so wie ich sie verstanden habe nicht drin wäre. Auf der
> Hauptseite z.B. alles, wenn man aber schon auf "Abteilungen" geklickt hat (und
> sich auf einer Unterseite befindet) könnte in einem dortigen Suchfeld nur noch der
> Basket Abteilungen durchscht werden - vom Admin so vorgegeben. "Shop" würde z.B.
> von da aus gar nicht zu wählen sein. Ob das Sinn macht ist sicherlich eine Frage,
> aber es wird vorkommen! Genau durch dieses Verhalten ist auch htaccess erklärt. Im
> öffentlichen "alles außer geschützten Bereich" durchsuche, ist man schon eingeloggt
> (htaccess) "alles" durchsuchen. So bekommt ein Nutzer der nicht eingeloggt ist
> gar nicht erst Suchergebnisse angezeigt auf die er garnicht zugreifen kann.

</htaccess issue>

- **Prio 55** add some nice feature found in other search engines: All these features should be able to be switched on /off by the admin in the configuration.
  - switch between short and long view of results (images “type 1“ and “type 2“)



Search results for 'improvement'

Match: All Format: Short Sort by: Score
Refine search: improvement Search

Documents 1 - 8 of 8 matches. More ☆'s indicate a better match.

SRD 3.5 > Legal > Legal Information☆☆☆☆☆

. . . derivative works and translations (including into other computer lang **improvement**, compilation, abridgment or other form in which an existing reproduce, license . . .

01/09/05, 9696 bytes

SRD 3.5 > Classes > Classes I☆☆☆

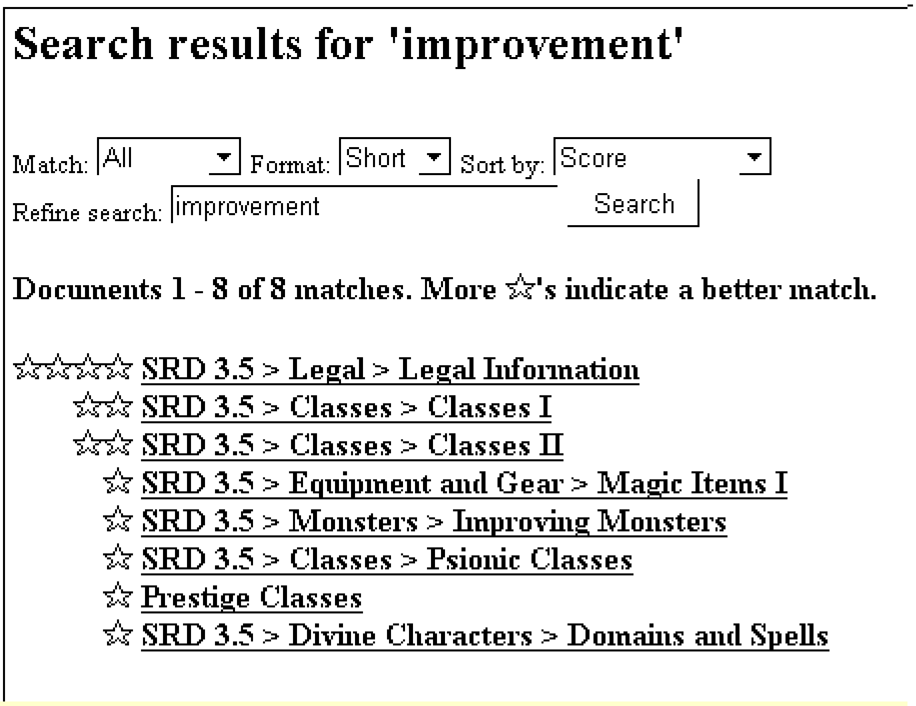
This material is Open Game Content, and is licensed for public use under tl BARBARIAN Alignment: Any nonlawful. Hit Die: d12. Class Skills The ba Craft (Int), Handle Animal . . .

01/09/05, 123362 bytes

SRD 3.5 > Classes > Classes II☆☆☆

This material is Open Game Content, and is licensed for public use under tl

type 1long



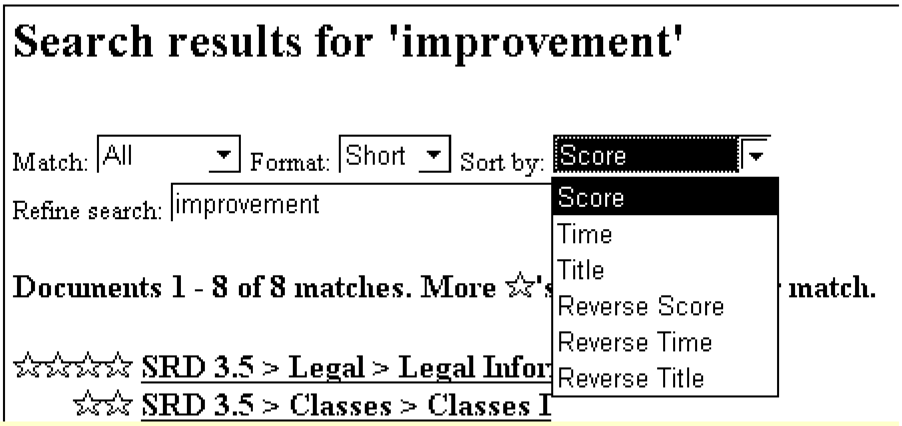
Search results for 'improvement'

Match: All Format: Short Sort by: Score
Refine search: improvement Search

Documents 1 - 8 of 8 matches. More ☆'s indicate a better match.

☆☆☆☆☆ **SRD 3.5 > Legal > Legal Information**
☆☆☆ **SRD 3.5 > Classes > Classes I**
☆☆☆ **SRD 3.5 > Classes > Classes II**
☆ **SRD 3.5 > Equipment and Gear > Magic Items I**
☆ **SRD 3.5 > Monsters > Improving Monsters**
☆ **SRD 3.5 > Classes > Psionic Classes**
☆ **Prestige Classes**
☆ **SRD 3.5 > Divine Characters > Domains and Spells**

type 2short



3Prio 75 different sorting to choose from + quick switch between boolean , TSEP and normal (same as the most search engine on the web) search

</ search>

< configuration>

- **Prio 55** add option to configuration: "Show search tipps when search returns no results". This will also output the search tipps in the search page (when using templates the admin can define the position) when a search returns no results. The "close this window" link needs to be removed in this case

</ configuration>

<HTML Weighting details – also see graphic below!>

Prio 66

better page weighting

Ideas and discussion about introducing better page weighting.
This is a result of the discussion between Manfred and Olaf as on Feb 9th 2005, 16:00

_*something* means this concerns the table called ...._*something* (usually tsep_something)

What this means:

We want to add a new way the pages are being ranked. As now, we count the occurrence of a word from the search term in the pages. The pages with the most hits has rank 1 (shown first in the results), the one with the least hits has the last rank (shown last in the results).
This is not really good.
We want to establish a ranking which pay attention to the HTML tags used in the page. Example a word in <h1> tag should have more weight to the ranking of the page A than the same word that shows up only in the body (<p>) tag of another page B.

Ideas:

The weights are set per basket definition: tsep_basket table will be similar to indexingprofiles

(iprofile) table. The _basket ID (id_basket) will be used at the tag-definitionen in tsep_internals table, field numtag (as makred (*) below)
The weights can be switched on and off per basket and globally in the configuration. If the weighting is off in configuration it will not be used in any baskets.

Tabelle tsep_internal

description	numericvalue	stringtag	numtag
h1	6	tagweight	(*)
h2	5	tagweight	(*)
h3	4	tagweight	(*)
b	3	tagweight	(*)
u	2	tagweight	(*)
i	1	tagweight	(*)
body	1	tagweight	(*)

We use the _internal table to add HTML tags to it. We use _internal because all handling structure has been developed for this already.
(*) will be used for identification of different baskets. (like stringtag 'indexer' and numtag id_profile). If a tag does not show up in the table (for example <p>) it has no special weight and will be counted only by other surrounding tags (at least by the 'body'-tag).

For the weighting it is important, that

```
<h1>großer<b>besonderer</b>Test</h1>
```

the word "besonderer" is weighted even higher than h1 because it is inside a <h1> and a <b> tag
To accomplish this the following text is placed in connection with a h1 tag: "großer besonderer Test". Additionally (!) "besonderer" is placed in connection with a <b> tag. Details how this is done are described below.
When searching the word "besonderer" is being weighted twice by the parser – once as <h1> and once as <b>. This way the word "besonderer" get more weight than being inside a <b> or inside a <h1> alone.

Indexing process:

When indexing we will look additionally for all tags in the document which are able to be weighted. This includes meta tags (from the <head> area). We will look in internal for the stringtag "tagweight" and save all values accordingly:

example:

```
<html>
	<meta name="keywords" content="Bernsteinstrasse, Kelten,
Germanen">
	<body>
	<h1>großer<b>besonderer</b>Test</h1>
	<p>ein großer test ist was tolles</p>
	</body>
	</html>
```

Table: **tsep_search**
id INTEGER(10) UNSIGNED NOT NULL AUTO_INCREMENT,
~~page_number INTEGER(4) NULL,~~ <=will be removed
protect_indexentry CHAR(1),
page_title VARCHAR(200) NULL,
page_url VARCHAR(200) NOT NULL,
page_file_size VARCHAR(10) NULL,
~~indexed_words LONGTEXT NULL,~~ <=will be removed
~~indexed_metawords TEXT NULL,~~ <=will be removed

Table: **tsep_searchtags**
id_searchtags
id_search
tag
text

id_searchtags	id_search	tag	text
Self explaining	Self explaining	meta keywords	Bernsteinstrasse, Kelten, Germanen
Self explaining	Self explaining	h1	großer besonderer Test
Self explaining	Self explaining	b	besonderer
Self explaining	Self explaining	body	großer besonderer Test ein großer test ist was tolles

<p> has no extra record because in this example p has no special weight – it does not show up in the tsep_internal table (not 0 – but no value at all: no table record for p!)

Searching:

We need to take care about baskets when searching.
internal information only:
The test for something belonging to a basket must take place (for speed reasons) directly in the sql command: Put a list of all “allowed” tsep_search/id's by usind a 'IN' in the sql
a) Basket 1 uses iprofile 3 and 4
b) iprofile 3 belongs to tsep_search 17 and 18
c) iprofile 4 belongs to tsep_search 13, 18 and 19
=> sql-Where also contains now: WHERE id IN (13, 17, 19)

_basket

_basket_iprofile

_iprofile

_iprofile_search

_search

_searchtag

Auswahl="1,2" (Jeder _basket-record enthält ein Prioritäts-Feld prio. Die höchste prio der *Auswahl* als **maxbasketprio** merken)

```
select idiprofile from ...
   where idbasket in (Auswahl)
result -> comma-sep-idiprofile-Liste
```

(GROUP muss sich auf Zielfelder beziehen, also noch ein s.url mit dem Select auswählen)

die url brauchen wir hier eigentlich gar nicht. Das wichtige ist die Sortierung nach total_weight... und als Resultat die idsearch'es

```
select idsearch from ...
   where idiprofile in (comma-sep-idiprofile-Liste)
result -> comma-sep-idsearch-Liste
```

```
select s.idsearch,sum(i.numericvalue) as total_weight_of_the_page
  from _search s, _searchtag st, _internal i
  group by s.url
  having MATCH (st.text) AGAINST ('$searchFor' $mysql_boolean)
     and (i.description = st.tag     and
          i.stringtag = 'tagweight' and
          i.numtag = :maxbasketprio)
     and (st.idsearch in (comma-sep-idsearch-Liste))
  order by total_weight_of_the_page DESC
result -> comma-sep-idsearch-Liste
```

```
for i = "anzeigen ab treffer x" bis
        "anzeigen ab treffer x" + anzahl anzuzeigende treffer je seite
   comma-sep-idsearch-Liste .= comma-sep-idsearch-Liste(i)
```

```
select * from _search s, _searchtag st
  where s.idsearch in (comma-sep-idsearch-Liste)
    and st.idsearch = s.idsearch and st.tag = 'body'
```

Navigation: ich weiß überhaupt nicht, wie ich's erklären soll, daher ein Beispiel:
Die Liste comma-sep-**idsearch**-Liste enthält bereits **alle** idsearch, die das **gesamte Suchergebnis** definieren - Beispiel
comma-sep-**idsearch**-Liste wäre 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20
Nimm an, der User möchte jeweils 7 Ergebnisse gleichzeitig anzeigt bekommen.
Dann wird beim ersten Durchlauf die for-Schleife mit "anzeigen ab treffer **1**" beginnen und 7 Durchläufe haben. Dadurch werden aus comma-sep-**idsearch**-Liste 1,2,3,4,5,6,7 angezeigt.
Es gibt ja jetzt noch 2 Folgeseiten (1.Seite 8-14 und 2.Seite 15-20)
Jetzt kommt der Clou: Es würde genügen, wenn wir bei den beiden Links zu den Folgeseiten (2 und 3) jeweils als Aufrufparameter die Liste der jeweils anzuzeigenden idsearch angeben. d.h.
der Link zu Seite 2 wäre dann zB "href='search.php?idlist=8,9,10,11,12,13,14'" und
der Link zu Seite 3 wäre dann zB "href='search.php?idlist=15,16,17,18,19,20'"
=> Dadurch ist es nicht mehr nötig nocheinmal zu Suchen!!! D.h. Die echte Suche nach dem, was der User finden will, wird nur ein einziges Mal durchgeführt!

Also die Lösung mit der Folgeseite find ich prinzipiell gut, ne, eigentlich super – aber stell dir mal vor der Suchende hat 1000 oder mehr Treffer – was dann? Gibt's dann nicht Probleme? Wir könnten die Zahlen auch in einer Variable, evt. hidden Formfield oder eben php intern oder einer temp Tabelle ablegen um sie beim Klick auf „Zeige Seite 500“ aufzulesen. So würden wir ggf. das Problem mit hinderten von Treffern umgehen.

Die Lösung ist sicher ein hidden field und via action=post aufrufen. Dadurch ist die Größe 1.egal und 2. steht auch die idsearch-Liste nicht im "Adresse"-Feld im Browser (was ziemlich unschön wäre).

Zum SQL: soweit so gut... bin nicht sicher warum du im HAVING description,. Stringtag und numbtag durch AND verknüpfst – liegt aber wohl an meiner inkompetenz ;-)

1. damit wir die Gewichtungen zum richtigen Basket aus _internal herauskriegen
2. viele würden sich wünschen, so "inkompetent" wie Du zu sein ;-)

</HTML Weighting details>

Manfred J's jobs

PRIO 60
<thumbnail web images>

- add thumbnail picture to pages

PRIO a

A default image can be set in configuration for all pages.
The administrator can define in configuration if the images for the pages should be shown

1. always (show default (can be empty!) picture or page picture
2. show page picture only (if no is set, show nothing)
3. show always no picture
4. show first picture of the html document which is at least the size of ..... (user defined)

PRIO b

MJ: Geniale Idee – das wollt' ich schon 'mal vorschlagen! Was ist da genau gemeint?
ON: gemeint: z.B. Screenshot von Page machen, als Thumbnail speichern und dann neben dem Ergebnis anzeigen (das war meine Idee). Bin drauf gekommen als ich http://lostgoggles.com/ gesehen habe (gibt's auch ein XPI für FireFox: http://docs.g-blog.net/code/mozilla_extensions/). Wenn der Screenshot automatisch erstellt werden könnte – noch besser. Dazu sollten die Bild Dateien aber direkt auf Platte gespeichert werden und nicht direkt in der Datenbank. Hier gibt es nur einen Namen (zufällig duch TSEP erstellt). Absolut genial wäre es natürlich wenn TSEP die „screenshots" beim indizieren automatisch erstellen könnte. Die Bilder sollten dann vielleicht auch bei der indexedit Seite erscheinen, austauschbar sein (=neues Bild hochladen) – evt mit TSEp richtig skalieren, ansonsten eben das nehmen das der Nutzer hochschickt und sich für einzelne Seiten abschalten lassen. Hierbei könnten noch viele Sachen einfließen. In der Configuration sollte dann ein Schalter sein für „Bilder zeigen (ja/nein)" und es sollte ein Default Bild geben was immer gezeigt wird wenn kein Bild vorhanden ist, Bilder jedoch gezeigt werden sollen. Zusätzlich könnte es noch ein Bild geben das bei „abgeschalteten Seiten" gezeigt wird. (das könnte ja das gleiche sein wie „kein Bild vorhanden")
So, das waren meine Ideen hierzu. Was meinst Du?

Look here:
http://www.aldostools.com/igrabber.html – shareware, can not crawl

http://www.tonec.com/products/wssh/index.html – shareware, activeX, can crawl

http://www.mozilla.org/projects/embedding/embedoverview/EmbeddingBasics.html – (probably too much for us)

http://dotnetjunkies.com/WebLog/sahilmalik/archive/2005/01/06/42130.aspx?Pending=true

http://studio.imagemagick.org/pipermail/magick-users/2005-January/014345.html

http://php-faq.de/q/q-grafik-webseite.html

http://www.cgi-interactive-uk.com/screen_capture_VB6_thumbnail_creation.html

http://www.sitepoint.com/forums/showthread.php?t=89660&page=4&pp=25 – good idea with script similar to this idea:

1. start new instance of webbrowser with desired URL
2. start screencapture program (i.e. irfanview). capture most front window

(browser with URL)
3. resize picture, save in specific place (i.e. c:\)
4. upload picture using php upload feature.

The following programs can automatically do the job (may not be free!)
web thummer: http://rcmedia.town-local.net/software/index.php?option=articles&task=viewarticle&artid=27

The administrator should be able to add more than one image per page! He should be able to add the <alt> tag description for each image as well.
</thumbnail web images>

< bastekts>

</introduce weighting of profiles in baskets>
Prio 45

admin should be able to apply weightings to profiles in baskets:

indexingprofile: Magazin, Onlineshop, Verkauf, Dienstleistungen

Basket "Find All" uses

indexingprofile: Magazin, Onlineshop, Verkauf
with weights: 2, 3, 1

The internal rank (not shown number) of a page is calculated by:
output position = (number of hits of a search term per page) * (the HIGHEST weight of all the profiles the pages belongs to).

The output order will be in the calulated order (output position).
</introduce weighting of profiles in baskets>

<profile baskets>
Prio 45

introduce “profilbaskets” which the admin can arrange and the user choose to which one (or more) the search should be limited

Idee (als mögliche spätere Erweiterung):
zB
/tsep/admin/
/tsep/include/
.
/tsep/templates/gruen/
/tsep/templates/gruen/verkauf/ <--------!!!!!!
/tsep/templates/gruen/onlineshop/ <--------!!!!!!
/tsep/templates/blau/
/tsep/templates/rot/

Diese Verzeichnisse können(!) bei den Basketdefinitionen ausgewählt werden. Ist nix angegeben, werden alle Templates aus zB /tsep/templates/gruen/ verwendet. Ist eines dieser Unterverzeichnisse ausgewählt, so wird beim Listen des Indexeintrages, der zu diesem Basket gehört, das "Indexeintraganzeige"-Template aus dem, dem

> Basket entsprechenden, Unterverzeichnis geholt!. Dadurch kann die Anzeige der Resultatseite unterschiedliche Zeilentypen enthalten, die optisch unterschiedlich sind.
>
> SChlage auf jedenfall vor das
> dann zur Pflicht zu machen zu einem Basket gehört ein Template - oder meinst du was dass Pflicht dann gar nich geht? Allerdings: (S.u. Basket in basket) - es wird immer das "höchste" Template des Basket Baums gewählt.

</profile baskets>

</ bastekts>

PRIO 70

\+ add ISPELL support (google: Ispell php, http://fmg-www.cs.ucla.edu/geoff/ispell.html, maybe even better: http://aspell.sourceforge.net/)

MJ: kenn ich nicht

ON: geht in die Richtung von dem „chevy“ = „chevrolet“ (oder wie man das auch schreibt). Damit meinte ich: Angenommen ich suche „cheyv“ (y und v vertauscht), dann werd ich z.B. gefragt „Oder meinen sie 'chevy'“ und kann direkt dadrauf klicken. Ist sowas wie bei google „Oder meinten Sie blablub“

===BELOW IS NOT IMPORTANT AT THIS TIME!===

+ have an username and password required to go to the admin area. (maybe in connection with the ADODB database work) -> this should be done by htaccess, so I, Olaf have moved this point down here.

needs testing:

+ catch (mysql) error messages (duplicate key) -> "insert IGNORE into" -> is this still needed ? (2004-12-02, ON)
+ to test: is IE buggy or our code - probably IE: search field extends suddenly when entering a special character like "õ"
+ test: indexstatus.php - does it always show the right number of pages found (to test: index many many pages, remove some pages, index again, see if number is correct)
+ to test: directory "x" could not be excluded from indeing in test
+ to test: does mysql <3.23 work? (-> no full text search)
+ test: functionn addQueryString($editFormAction) in mmexfunctions.php - do we need this?
+ make markup of "some words" correct - the individual words are marked, but even if they are not like in the quotes

should be solved:

+ little bug: when indexing the first time: "0 files found" is shown
+ make tsep config variables global / use sessions (?not sure if I want to do this)
+ indexedit overview -> go to details -> go back .. go back to the same page: use url parameters ?...
+ pagenavigation.php -> need this because somewhere here is an error wich might make counting mistakes: take 10 resultes/page, go the second last page, switch to 100 results - wrong count. Fixed for now with a workaround in file.

kind of solved:

Done (see also history.txt):

- ~~done for 0.940:~~
- ~~add a hint to words which are in more than 50% of the searches might not return anything!~~
- ~~documentation: add hint: 3 character are mysql restriction~~
- ~~documentation: make clearer what quick.htm and complete.htm are all about~~
- ~~documentation: add hint: search from indexer is not logged~~
- ~~documentation: hint: look at tsepsearch.php for details how to build your own search page~~
- ~~add a hint in the docs that password protected directories will not be indexed (correctly). (<= is this true when we use fopen???) even better would be to access .htacces protected directories where the user would need to enter password before the indexing starts~~
- ~~done in version 0.938:~~
-
- ~~PRIO 1~~
- ~~+ check excluded directories in the section of code that recursively calls search_get instead of the section that parses the file? This would short-circuit the unneceessary recursive calls. (SF bug 1082257)~~
-
- ~~PRIO 2~~
- ~~+ correct the number of pages shown after creating a new index~~
- ~~MJ?~~
- ~~ON: Sollte kein Problem sein (=einfach – der Wert darf eben erst nach dem indizieren durch das select count from bla erstellt und angezeigt werden)~~
-

- ~~PRIO 3~~
- ~~+ grab TSEP path for the installation and the indexer.php (by clicking on a link next to the field)~~
- ~~MJ: was ist damit gemeint?~~
- ~~ON: im configuration „TSEP Pfad" -> hier kann automatisch der richtige eingetragen werden $_server[path translated] oder so~~
-
- ~~PRIO 4~~
- ~~+ add extra field "meta" to _search table where meta tags are stored. This will enable to specifically output meta tags/search only in meta tags/put higher value to meta tags~~
-
- ~~PRIO 5~~
- ~~+ Profiles (hat in der Todo gefehlt, daher doku hier zusammengefasst)~~
- ~~wir sollten es IndexProfiles oder (besser vielleicht) IndexCategories nennen, weil ich mir vorstellen kann, daß das dann einmal auch in der Benutzeroberfläche eingebaut wird, wo der User dann auch nur innerhalb einer Kategorie suchen kann.~~
- ~~In der Benutzeroberfläche sollte das nur optional angezeigt werden (optional aber definiert über eine Einstellung via configuration.php!).~~
-
- ~~Ist: Momentan wird vor Indizierungsstart der gesamte Index gelöscht.~~
- ~~Soll: Um zu Indizieren, muß ein Profil ausgewählt sein (zumindest das default-Profil).Die beim indizieren ausgewählte Profile-Id (idprofiles), muß auch im tsep_search-Record (=Indexeintrag) abgespeichert werden.~~
- ~~Dadurch kann identifiziert werden, welches Profile für die Indizierung dieser Seite "verantwortlich" war.~~
- ~~Wir nun neuerlich indiziert, so werden zu Indizierungsbeginn nicht ALLE tsep_search-Records gelöscht, sondern nur jene mit der ausgewählten ProfileId (mit Ausnahme jener, die das Sperrflag gesetzt haben).~~
- ~~Profileverwaltung: Liste der Profile + Anzahl zugehöriger Indizes, Löschen, Umbenennen. Neue Profile werden im Indexer-Startschirm angelegt (mit SaveAs-Button).~~
- ~~auf der indexoverview Seite: Profil (oder "Alle Profile") wählen, mit einem anderen Schalter sagen nur gesperrte / entsperrte / alle Zeigen.~~
- ~~Es muss auch eine neue Tabelle her "Profiles".~~
- ~~Felder:~~
- ~~idprofiles (zahl, autoincrement, primärschlüssel)~~
- ~~profilename (text 50, not null, unique index)~~
- ~~In dem neuen Feld der Tabelle internal wird dann natürlich nur der idindexingprofiles gespeichert. (Normalisierung!) Statt~~
- ~~9,defaultprofile,listFilenamesOnly~~
- ~~10,defaultprofile,ext_include~~
- ~~11,onlineshop,XdirName~~
- ~~12,onlineshop,Xwebdir~~
- ~~also:~~
- ~~Tabelle~~
- ~~Profiles:~~
- ~~1,default~~
- ~~2,onlineshop~~
- ~~Internal:~~
- ~~9,1,listFilenamesOnly~~
- ~~10,1,ext_include~~
- ~~11,2,XdirName~~
- ~~12,2,Xwebdir~~

~~<how to use stopwordlists> added 2005-01-31, by Olaf Noehring~~

~~a) Add the possibility to include a txt file with stopwords OR~~

~~b) read a files into the stopwords list OR~~
~~c) copy + paste a list of words, either seperated by "," or by " " (space) into the existing stopwords page.~~
~~Most probably we will do c).~~

- ~~</how to use stopwordlists>~~
- ~~translation script: online translation of the strings.~~
  ~~People can apply to get a login/pwd from us (TSEP programmers).~~
  ~~After they log in they choose which language they want to translate.~~
  ~~We have to allow who can translate which language. Then the user (translator) can choose a source language (english default, all mising string in other languages will be substitued by their english equivalent) The user can change other translations. Missing translations will be shown in a different color. On the bottom of the page the translator can pres "submit" and will submit his translations. A file should be created "language-n.php" in the folder of the language. The "-nnn" is a number. The number will enable us to keep older versions. This file contains the translations of all the strings of that language. In an overview page in the beginning we can show how many translations are missing of that language. The translation page should be displayed in the ISO characters set which has been set for a specific language or ISO for english language if nothing is specified. Examples:~~